

УДК 164.07

DOI: 10.17223/1998863X/40/10

В.В. Целищев

## ИНТЕНСИОНАЛЬНОСТЬ ВТОРОЙ ТЕОРЕМЫ ГЕДЕЛЯ О НЕПОЛНОТЕ<sup>1</sup>

*Рассмотрены отличия Второй теоремы Геделя о неполноте от Первой теоремы, имеющие первостепенное значение для возможной реабилитации Программы Гильберта. Доказано, что сам Гедель не усматривал во Второй теореме непосредственную угрозу этой программе, что Вторая теорема не может быть интерпретирована в качестве столь же радикального приговора для программ в основаниях математики, как в случае Первой теоремы. Суть отличия заключается в том, что если Первая теорема является экстенсией, то Вторая теорема – интенсией. Оспаривается формализация концепции непротиворечивости, используемая в доказательстве Второй теоремы, с представлением других способов выражения свойства непротиворечивости, для которых Вторая теорема не имеет места. Исследуется, что можно считать формальным выражением содержательной концепции непротиворечивости.*

Ключевые слова: интенциональность, теорема Геделя, самореференция, непротиворечивость, доказательство.

Доказательство К. Геделем Первой теоремы о неполноте арифметики в статье 1931 г. связывают с крушением двух важнейших программ в основаниях математики, логицизма Б. Рассела и формализма Д. Гильберта. Кодификация математического мышления в системе *Principia Mathematica* Б. Рассела и А.Н. Уайтхеда оказалась неполной. Построение в достаточно богатых формальных системах арифметики так называемого геделева неразрешимого предложения, истинного, но не доказуемого в этой системе, показало неполноту *Principia Mathematica* и родственных ей систем. Другими словами, не все содержательно истинные утверждения арифметики могли быть доказаны, что в известной степени подрывало саму идею полной формализации математики. Поскольку среди таких неразрешимых предложений оказалось утверждение о непротиворечивости системы, Программу Гильберта доказательства непротиворечивости арифметики средствами самой арифметики большинство исследователей посчитали неосуществимой. Строгий результат, утверждающий невозможность такого доказательства, представлен Второй теоремой Геделя о неполноте.

Введение нового способа обсуждения формальных систем арифметики средствами самой формальной системы с помощью кодирования синтаксических структур арифметическими средствами, так называемая геделева нумерация, и связанные с этим тонкие технические детали осложнили понимание функционирования и роли этих теорем, породив тем самым среди философов значительное число искаженных представлений и просто заблуждений в их интерпретации. Так, среди распространенных мифов о теоремах Геделя не

---

<sup>1</sup> Статья подготовлена при поддержке Российского научного фонда, проект 16-18-10359.

последнее место занимает убеждение, что Вторая теорема является чуть ли не тривиальным следствием Первой теоремы. Но сам Гедель не стал доказывать Вторую теорему, сделав лишь несколько замечаний о возможности такого доказательства. Оно появилось во втором томе «Оснований математики» Гильберта и Бернаиса [1].

Помимо исторических деталей о разном происхождении доказательств двух теорем Геделя о неполноте, важно и различие их характера. Эта тема долгое время вообще не обсуждалась и до сих пор является относительно неизвестной [2]. Между тем обнаружение отличия Второй теоремы от Первой имеет первостепенное значение для возможной реабилитации Программы Гильберта. При этом нужно учесть, что сам Гедель не усматривал во Второй теореме непосредственную угрозу этой программе. Г. Крайзель писал:

Как подчеркивал сам Гедель еще в 1931 г. ...его Вторая теорема о неполноте не имеет отношения ни к какой разумной постановке вопроса о непротиворечивости. Ведь если  $\text{Con}_T$  сомнительна, то почему ее надо доказывать в  $T$  (а не в системе, несравнимой с  $T$ )? [3. Р. 48].

Прозрения Геделя по этому поводу не были подкреплены какими-либо аргументами в то время, и только спустя три десятилетия С. Феферман озвучил причину того, почему Вторая теорема не может быть интерпретирована столь же радикальным приговором для программ в основаниях математики, как Первая теорема в отношении логицизма. Суть отличия заключается в том, что если Первая теорема, согласно терминологии С. Фефермана, является экстенсией, то Вторая теорема – интенсией [4]. Причины относительной неизвестности этой проблематики таковы. Как известно, популярная интерпретация геделева предложения (инициированная, кстати, самим Геделем, и в настоящее время квалифицируемая как эвристическое замечание) состоит в том, что это предложение говорит о самом себе, что оно не доказуемо, и поскольку оно на самом деле недоказуемо, оно оказывается истинным. Другими словами, здесь имеет место самореференция, известная еще со времен парадокса «Лжеца». Как свидетельствует К. Смори́нски, одной из причин слабого интереса математиков к проблематике самореференции в связи с геделевскими теоремами о неполноте является их «замешательство» перед вопиющей интенциональностью этих теорем [5. Р. 110]. На самом деле, речь должна идти о своеобразии и необычности Второй теоремы: дело в том, что семантика математического дискурса характеризуется экстенциональностью теоретико-модельного рода. Редкие исследования посвящены интенциональным аспектам математики [6], а интенциональная семантика *Principia Mathematica* была отвергнута в пользу экстенциональной аксиоматической теории множеств Цермело.

Оживление интереса к интенциональному аспекту Второй теоремы было связано с сомнениями ряда исследователей в давно установленном вердикте о провале Программы Гильберта [7]. Эти сомнения мотивированы двумя обстоятельствами. Во-первых, демонстрируется, что методология финитарной арифметики не может быть схвачена формальной системой типа Арифметики Пеано. Во-вторых, оспаривается формализация концепции непротиворечивости, используемая в доказательстве Второй теоремы, поскольку есть другие

способы выражения свойства непротиворечивости, которые не подлежат этой теореме. В данной статье будет исследовано, что можно считать формальным выражением содержательной концепции непротиворечивости.

Феферман отметил специфику получения предложения во Второй теореме. Детлефен [8] и Ауэрбах [9] обратили внимание на интенциональный характер предиката доказуемости в конструировании геделевого неразрешимого предложения. Предтечей такого понимания был Г. Крейзель, который подчеркнул при исследовании проблемы Генкина о выражении предложения о самодоказуемости (в некотором смысле дуального к геделевскому предложению), что важнейшее значение имеет то, как выражается доказуемость и как конструируется формула для выражения собственной доказуемости предложения [10].

Интенциональный характер Второй теоремы является проявлением феномена самореферентности языка. Поскольку сама по себе тема самореферентности далеко выходит за пределы собственно интерпретации теорем о неполноте, разумно ограничиться только теми аспектами интенциональности математического дискурса, которые связаны с самореферентными конструкциями геделевского типа. Последние представляют собой арифметические предложения, приписывающие себе определенное свойство, в частности доказуемость или недоказуемость. Ф. Холбах и А. Виссер [11] называют три стадии вовлечения их в интенциональность: во-первых, это кодирование числами выражений языка, или геделевская нумерация синтаксических образований. Во-вторых, это определение формулы, которая выражает это свойство. В-третьих, конструирование из этой формулы самореферентного предложения. Помимо выбора различных способов кодирования, которые воздействуют на остальные две стадии, требуется тщательный выбор представляющей формулы, выражающей свойство. Именно эти обстоятельства являются источниками интенциональности. В более общей постановке вопроса, дальнейшими факторами интенциональности являются выбранный язык и формализация. Наконец, важнейшим фактором является вопрос о соотношении содержания математического результата, выраженного на обычном языке, и формализации этого результата. Все эти факторы создают «ауру» интенциональности, которая делает Вторую теорему о неполноте Геделя столь отличной от Первой его теоремы.

Относительное невнимание к этой стороне геделевских теорем К. Смори́нски увязывает с расхождением интересов философски настроенных логиков и математиков, заинтересованных в технических аспектах самореферентности [5. Р. 111]. Это расхождение свидетельствует, с его точки зрения, о подлинной смене парадигм в плане соотношения логики и философии. Постепенное вытеснение философских проблем из повестки дня математических логиков видно из того, что изначальная мотивация – устранение парадоксов – сменилась на установку поиска новых технических средств в теории доказательства. Преобладание аксиоматического подхода в качестве показателя окончательного развода начинается, по свидетельству Смори́нски, с 1970-х. Это обстоятельство хорошо просматривается на отношении к теоремам Геделя о неполноте, доказательство которых в настоящее время связывается, в отличие от оригинального геделевского, с Диагональной Леммой, или

Теоремой о Неподвижных Точках, а также с теоремой Леба. Соответствующая трансформация происходит в понимании природы самореферентности геделевых предложений, вплоть до полного отказа им в таком статусе.

В строгой трактовке самореферентный характер можно усмотреть из Диагональной Леммы, согласно которой для произвольного рекурсивного непротиворечивого расширения  $S$  формальной системы  $Q$  (Арифметика Робинсона) можно найти такое предложение  $G$ , что в  $Q$  выводимо

$$G \leftrightarrow \neg \text{Pr}_S(\ulcorner G \urcorner),$$

где  $\text{Pr}_S^*(y) = (\exists x) \text{Proof}_S(x, y)$ , в свою очередь,  $\text{Proof}_S(x, y)$  есть  $\Delta_0$  формула, определяющая  $\text{Proof}_S$  в  $\omega$ , и теперь уже  $\text{Proof}_S = \{ (m, n) \in \omega^2 : n = \ulcorner \varphi \urcorner \text{ для некоторой формулы } \varphi \text{ и } m \text{ есть геделево число доказательства } \varphi \text{ в } S \}$ . (Здесь  $\omega$  есть множество натуральных чисел, а через  $\ulcorner \varphi \urcorner$  обозначается геделев номер формулы  $\varphi$ .)

Предложение  $G$  оказывается доказуемо эквивалентным арифметическому предложению

$$(\forall x) \neg \text{Proof}_S(x, \ulcorner G \urcorner).$$

Гедель в своем оригинальном доказательстве использовал именно арифметический предикат. Однако Р. Карнап, который одним из первых прочитал рукопись Геделя, заметил общий характер утверждения  $G \leftrightarrow (\forall x) \neg \text{Proof}_S(x, \ulcorner G \urcorner)$ , в том смысле, что нет прямой зависимости от предиката доказуемости, и предложил более общий вариант, согласно которому в  $Q$  доказуемо

$$\varphi \leftrightarrow \alpha(\ulcorner \varphi \urcorner)$$

с формулой  $\alpha$ , заменяющей предикат истины. Самореферентность возникает ввиду того, что формула  $\varphi$  отображается в предложение, эквивалентное ему в  $Q$ . Более точно, Диагональная Лемма, или Теорема о Неподвижной Точке, утверждает, что имеется некоторое предложение такое, что мы можем доказать в системе  $Q$ , что это предложение эквивалентно утверждению о геделевом номере этого самого предложения. Другими словами, предложение говорит нечто о своем собственном геделевом номере. Но что именно? Поскольку Теорема (или Лемма), согласно которой  $\varphi \leftrightarrow \alpha(\ulcorner \varphi \urcorner)$  применима к любому предложению  $\alpha$ , можно найти предложение такое, которое говорит о своем собственном геделевом номере все что угодно. Суть геделевского доказательства заключается в том, чтобы тщательно выбрать то, что оно *должно* сказать о своем геделевом номере. Гедель выбрал свойство нетеоремности, выраженное через отрицание предиката доказуемости [12. Р. 190]. Искомое предложение в доказательстве Геделя и есть геделево неразрешимое предложение  $G$ .

Выбор в качестве  $\alpha$  свойства «нетеоремности» вносит свой вклад в традиционное понимание  $G$  как утверждения, говорящего о собственной недоказуемости. Другими словами,  $G$ , в силу Диагональной Леммы, является в некотором смысле самореферентным. Однако есть сомнения в том, что в случае  $G$  мы имеем дело с подлинной самореферентцией. Действительно, Диагональная Лемма представляет эквивалентность в терминах истинностных функций. Но такой эквивалентности вряд ли достаточно для обоснования самореферентности. Б. Булдрт уподобляет ситуацию предполагаемой самореферентности с обычной эквивалентностью в арифметике:  $1+1 = 2$  и  $2 \times 3 = 6$  эквива-

лентны в терминах истинностных функций, но в установлении такой эквивалентности нет ничего, что «говорило» бы о значении этого факта. Другими словами, установление эквивалентности в терминах истинностных функций не дает эквивалентности значений. Булдрт полагает, что сама по себе Диагональная Лемма, или Теорема о Неподвижной Точке вида  $\varphi \leftrightarrow \alpha(\ulcorner \varphi \urcorner)$ , не выделяет именно геделева предложения в качестве  $\varphi$ , поскольку таких неподвижных точек может быть много. Это является результатом использования экстенциональной логики, в которой основное значение имеют лишь истинностные значения [13].

Ситуация с установлением значения геделевого предложения, которое «говорит само о себе, что он не доказуемо», имеет дело и с более общими соображениями. Можно ли обосновывать истинность  $G$  на факте самореферентности  $G$ ? Коль скоро в системе  $Q$  мы имеем выводимое выражение  $G \leftrightarrow \neg \text{Pr}_S(\ulcorner G \urcorner)$ , а вторая часть эквивалентности равносильна  $(\forall x) \neg \text{Proof}_S(x, \ulcorner G \urcorner)$ , все три выражения являются взаимозаменяемыми. Но что это значит в содержательном смысле? Последнее является арифметическим утверждением, а  $G$  является метаматематическим артефактом. Какой смысл заключается в их эквивалентности? На этот вопрос можно отвечать по-разному. Например, Г. Серени полагает, что подлинным смыслом обладают метаматематические операции, а арифметический предикат, имеющий дело с огромными в результате геделевой нумерации числами, просто непостижим [14]. Другая реакция заключается в обращении к интенциональным аспектам геделева предложения.

Как уже отмечалось выше, на важность интенциональных аспектов теорем о неполноте Геделя указал Г. Крайзель [15], который, в частности, заметил, что есть различие между предложениями, выражающими определенные свойства самих себя, и просто неподвижными точками. Если Гедель сконструировал предложение  $\varphi$  такое, что

$$\varphi \leftrightarrow \neg \text{Pr}(\ulcorner \varphi \urcorner),$$

где  $\text{Pr}$  – предикат доказуемости, то Л. Генкин [16. Р. 160] поставил вопрос, который при обобщении принимает следующий вид: пусть имеется предложение  $\varphi$  такое, что оно удовлетворяет утверждению

$$\varphi \leftrightarrow \text{Pr}(\ulcorner \varphi \urcorner).$$

Является ли предложение  $\varphi$  доказуемым?

К. Смори́нски так комментирует связь проблемы Генкина с необходимостью учета интенционального аспекта проблемы:

...тут есть философский вопрос: Отношение между формулой  $\alpha$  [в выражении  $\varphi \leftrightarrow \alpha(\ulcorner \varphi \urcorner)$ ] и неподвижной точкой  $\varphi$ , сконструированной методом Геделя, сильнее, чем эквивалентность  $\varphi$  и  $\alpha(\ulcorner \varphi \urcorner)$ ;  $\varphi$  буквально имеет форму  $\alpha(t)$  для некоторого термина  $t$ , который находит численное значение коду  $\ulcorner \varphi(t) \urcorner$ . Таким образом,  $\varphi$  буквально выражает  $\alpha(\ulcorner \varphi \urcorner)$ . Генкин не хотел знать, не является ли некоторое предложение, случайно эквивалентное утверждению о его собственной доказуемости, доказуемым...он хотел узнать, *выражает* ли предложение, *сконструированное* таким образом, чтобы показать свою собственную доказуемость, доказуемым [5. Р. 114].

Два обстоятельства тут важны: во-первых, было отмечено, что устанавливаемая Теоремой о неподвижных точках эквивалентность формулы  $\alpha$  и неподвижной точки  $\varphi$ , а именно  $\varphi \leftrightarrow \alpha[\varphi]$ , на самом деле выходит за пределы экстенциональной эквивалентности, поскольку  $\varphi$  буквально выражает  $\alpha[\varphi]$ . Стало ясно, что есть различие между предложениями, которые выражают некоторые свойства самих себя, и просто неподвижными точками, т.е. различие между экстенциональной и интенциональной трактовками геделевых предложений.

Другими словами, формальная эквивалентность утверждений, устанавливаемая Теоремой о неподвижных точках, не является достаточной для выражения интенционального аспекта, т.е. для понимания того, что выражается соответствующей формулой. Уже одного этого достаточно для понимания приведенного выше замечания К. Сморинского о «вопиющей» интенциональности теорем Геделя. Однако «степень» интенциональности Второй теоремы намного выше, чем у Первой.

Первая теорема представляет собой условное утверждение, антецедент которого есть утверждение о непротиворечивости системы, а консеквент – формальное предложение о недоказуемости геделева предложения. Способ доказательства Первой теоремы таков, что этот самый консеквент эквивалентен геделеву предложению. Отсюда легко заключить, что при условии непротиворечивости системы формализованное утверждение непротиворечивости не является выводимым. Это как будто следствие Первой теоремы фигурирует в качестве наброска доказательства Второй теоремы, и оно должно было превратиться в строгое доказательство через конструирование формального вывода. Но именно в этом понятии формального вывода заключаются концептуальные трудности. Что, собственно, является невыводимым согласно этой теореме? Это формула, которая по предположению *говорит*, что формализм непротиворечив. Но это уже семантический дискурс; другими словами, уже «строгому» варианту Второй теоремы присущ семантический, или интенциональный, характер.

Действительно, согласно Первой теореме, речь идет о предикции простого синтаксического свойства неполноты членам обширного класса формальных систем. А вот в случае Второй теоремы речь идет о предикции более сложного свойства, традиционно обозначаемого через  $\text{Con}$ . Между тем остается неясным соотношение неразрешимого предложения из Первой теоремы  $G$  и утверждения из Второй теоремы  $\text{Con}$ , несмотря на то, что в рамках теории выводимо утверждение об экстенциональной эквивалентности  $\text{Con} \leftrightarrow G$ . Эта эквивалентность подсказывает нам, что точно так же, как  $G$  говорит о своей собственной недоказуемости,  $\text{Con}$  должно говорить о непротиворечивости формализма. Но конструирование  $G$  с помощью арифметического предиката доказуемости не дает оснований для подобного вывода. Экстенциональность в Первой теореме проявляется в том, что ограничения на арифметический предикат, используемый для конструирования  $G$ , выражает свойство *есть доказательство (чего-то)*. Но «требование экстенционального согласия в отношении предиката доказуемости, в общем случае, не дает, однако, экстенционального согласия в отношении утверждения о непротиворечивости. То есть для некоторой теории  $T$  и некоторой фор-

мулы  $\text{Con}_T$  ложно: « $\text{Con}_T$  истинно, если и только если  $T$  непротиворечива» [9. Р. 342]. Мораль, выводимая из этих соображений, состоит в необходимости учитывать интенциональные аспекты соотношения технических деталей и семантических вопросов при интерпретации Второй теоремы Геделя о неполноте. Ауэрбах упоминает в этой связи цитированную выше работу С. Фермана, который показал, что прос той нумерической правильности условий формальной выводимости Гильберта – Бернайса недостаточно в том отношении, что они неадекватны для определенных результатов. (Адекватные условия вызывают к экстенциональности, а неадекватные – к интенциональности) [4].

Особенность этой теоремы, редкая в математике, состоит в вовлечении значения или интенции некоторого математического символизма, в противоположность просто указанию или экстенционала. Это проявляется в условии, что  $\text{Con}_T$  *выражает*  $T$ -невыводимость непротиворечивости  $T$ . Это не дополнительное условие, добавленное ради философии. Если кандидат на роль предложения о непротиворечивости не выражает непротиворечивости  $T$ , если он сконструирован из предиката, который представляет, но не выражает « $x$  есть  $T$ -выведение  $y$ », аргумент о Второй теореме не может гарантировать доказательства для предложения, поскольку предикат не удовлетворяет условиям выводимости, о которых речь пойдет ниже. Даже если можно показать, что это предложение недоказуемо в  $T$ , трудно видеть, какое значение будет иметь этот факт, поскольку это предложение не выражает непротиворечивости  $T$ . В противоположность этому невыводимость в  $T$  предложения, которое все-таки выражает, что  $T$  непротиворечиво, значимо, поскольку с первого взгляда угрожает Программе Гильберта.

Для понимания существенного различия в характере обеих теорем нужно иметь в виду три «ипостаси» каждой из них. Во-первых, это формулировка результата в обыденном языке, в которой излагается результат, полученный при доказательстве теоремы. Во-вторых, это представление теоремы в формальном языке. Наконец, это представление доказуемости этой теоремы в конкретной формальной системе. При переходе от первой «ипостаси» ко второй лежат значительные трудности, на что обратил особое внимание Д. Ауэрбах [9]. В частности, формализация доказательства Первой теоремы производится Геделем в формальной системе *Principia Mathematica* (PM), в то время как аналогичная формализация в случае Второй теоремы, которая говорит о недоказуемости непротиворечивости Арифметики Пеано (PA) в самой PA, производится в PA, которая имеет меньшие выразительные возможности по сравнению с PM. В «приспособлении» к PA лежат значительные технические трудности, которые являются одной из причин того, что доказательство Второй теоремы редко появляется в учебниках.

Здесь ключевую роль играют два фактора. Во-первых, что именно выбрать в качестве формализации утверждения «(теория)  $T$  непротиворечива». Стандартный выбор в виде геделевского  $\text{Con}_T$  хорош, поскольку  $\text{Con}_T$  действительно не доказуемо.

Предикат доказуемости для предложения  $\phi$  в теории  $T$  обозначается через  $\text{Prov}_T(\ulcorner \phi \urcorner)$ , где  $\ulcorner \phi \urcorner$  есть геделев номер предложения  $\phi$ . Этот предикат является ключевым понятием в представлении последней «ипостаси» Первой теоремы « $\text{PA} \propto G \leftrightarrow \neg \text{Prov}(\ulcorner G \urcorner)$ », где  $G$  – геделево неразрешимое предло-

жение. Далее, в первопорядковой логике РА используется константа « $\perp$ », интерпретируемая как ложь. Центральным понятием Второй теоремы является непротиворечивость формальной системы, которая в введенной нотации представлена формулой  $\text{Con} =_{\text{Df}} \neg \text{Prov}(\ulcorner \perp \urcorner)$  (недоказуемость ложного утверждения). Это одна из экспликаций понятия непротиворечивости, но это не означает, что не может быть других формул для утверждения о непротиворечивости, которые доказуемы в Т. Нестандартный выбор означает, что Вторая теорема не является прямым опровержением Программы Гильберта. Эта точка зрения противоречит укоренившейся традиции в интерпретации теорем Геделя. Г. Крайзель заявляет:

Ближе к концу [статьи о формальной неполноте] он [Гедель] предупреждает против соотнесения Второй теоремы и Программы Гильберта... [17. Р. 125].

Нестандартные концепции доказуемости непротиворечивости теории внутри самой теории, например на основе предиката доказуемости Россера, представляют серьезную проблему, которая будет обсуждена ниже. Сейчас же имеет смысл добавить, что нестандартные концепции не удовлетворяют условиям выводимости Гильберта – Бернайса, соблюдение которых необходимо для доказательства Второй теоремы. Это предвидел сам Гедель. Если в теории Т имеется формально неразрешимая формула, то мы можем без противоречия присоединить к Т ее отрицание. Это означает нарушение одного из условий выводимости.

Таким образом,  $\text{Con}_T$  не гарантирует [условия выводимости]... Гедель высказал это предостережение в работе [18] со всей возможной ясностью еще до открытия неполноты. Однако [18] произвела гораздо меньшее впечатление, чем следующие банальности... В более поздних «популярных» произведениях Гедель всегда оказывал таким банальностям надлежащее почтение... Первоначальное название статьи... обобщающей [знаменитую статью в журнале *Dialectica*] не содержало слова «непротиворечивость»; оно было добавлено в посмертной публикации по настоянию Геделя. Он слишком хорошо знал рекламную цену этого модного слова, которое, вопреки его собственному взгляду на предмет, сделало его Вторую теорему о неполноте более популярной, чем Первая теорема [3. Р. 46, 49].

Таким образом, наличие нестандартных формализаций концепции непротиворечивости поднимает проблемы, которые отчасти совпадают с идеями самого Геделя.

Вторым фактором явился, по выражению Сморинского, парадигмальный сдвиг в теории доказательства при использовании модальной логики, в частности, использование модального оператора « $\Box$ », интерпретируемого как «доказуемо» в применении к формулам [5]. Концептуальное и нотационное улучшение при замене выражения  $\text{Prov}_T(\ulcorner \varphi \urcorner)$  выражением  $\Box_T(\varphi)$  заключается в двойной функции знака « $\Box$ »: во-первых, это сокращение для  $\text{Prov}$  и, во-вторых, поглощение «скобок», превращающих правильно построенную формулу  $\varphi$  в стандартное число, представляющее геделев номер предложения  $\varphi$ .

Коль скоро имеется формальное выражение непротиворечивости  $\text{Con}$ , три «ипостаси» Первой теоремы предстают в следующем виде:



1. Если РА непротиворечива, тогда в РА недоказуемо геделево предложение G.

2.  $\text{Con} \rightarrow \neg \Box G$ .

3.  $\text{PA} \Box \text{Con} \rightarrow \neg \Box G$ .

Три «ипостаси» Второй теоремы выглядят так:

1. Если теория Т непротиворечива, тогда в Т не может быть доказана ее непротиворечивость.

2.  $\text{Con} \rightarrow \neg \Box \text{Con}$ .

3.  $\text{T} \Box \text{Con} \rightarrow \neg \Box \text{Con}$ .

К теории Т предъявляются определенные стандартные требования, на которых мы здесь не останавливаемся. Что касается предиката доказуемости,  $\text{Prov}(\varphi)$  или  $\Box_T(\varphi)$ , то его выбор должен быть осторожным. Как известно, он регулируется условиями выводимости Гильберта – Бернайса: для любых предложений  $\varphi$  и  $\psi$ :

(C1) Если  $\text{T} \Box \varphi$ , тогда  $\text{T} \Box \Box \varphi$ ,

(C2)  $\text{T} \Box \Box (\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi)$ ,

(C3)  $\text{T} \Box \Box \varphi \rightarrow \Box \Box \varphi$ .

Концепция Con интерпретируется здесь как невозможность вывода ложного утверждения. Таким образом, выражение концепции непротиворечивости является стандартным для большей части изложений сути Второй теоремы. Однако Con может быть не единственным формальным выражением концепции непротиворечивости как невозможности вывода ложного утверждения. Речь идет о поиске какого-то другого предложения, устанавливающего непротиворечивость, скажем, РА. П. Смит приводит примеры имеющих в литературе двух явных альтернатив Con, которые он обозначает как  $\text{Con}'$  и  $\text{Con}''$ , демонстрируя, что все три версии попросту эквивалентны. Ввиду такого результата он называет Con каноническим утверждением непротиворечивости РА [19. Р. 238]. Важный пример подобного рода привел А. Мостовский [20].

Это относится к очень четкой неформальной концепции непротиворечивости системы, при которой невозможно вывести ложное утверждение. Принимая Con в качестве канонического выражения непротиворечивости, Вторая теорема формулируется в переводе на обыденный язык следующим образом: если в достаточно богатой теории доказуемо каноническое предложение о непротиворечивости, тогда эта теория противоречива. Естественно, что термин «доказуемо» в предыдущем предложении заслуживает кавычек.

А если мы попытаемся найти неканоническое предложение для непротиворечивости, скажем,  $\text{Con}^*$ , удовлетворяющее условию доказуемости в РА? Для этой цели можно попробовать слегка изменить выражение для  $\text{Prf}$ , добавив в его определение явное упоминание о невозможности получения « $\perp$ », например

$$\text{Prf}^*(x, y) = \text{Df } \text{Prf}(x, y) \ \& \ (\forall v < x) \neg \text{Prf}(v, \lceil \perp \rceil).$$

Легко показать, что  $\text{Prf}$  представляет интуитивное отношение геделевых чисел доказуемого утверждения и его доказательства [19. Р. 262]. Однако при таком определении  $\text{Prf}^*$  и, соответственно, определимого  $\text{Prov}^*$  можно

доказать теорему, согласно которой в системе  $PA$  выводимо  $Con^*$ , или  $PA \propto Con^*$ , что и было целью нашего поиска.

Поскольку в  $T$  «доказана» ее собственная непротиворечивость в варианте  $Con^*$ , ясно, что новая концепция в чем-то нарушает условия, накладываемые на предикат доказуемости. И в самом деле, можно показать, что приведенное выше определение  $Prf^*$  и, стало быть,  $Prov^*$  не удовлетворяет условиям Гильберта – Бернаиса; в частности, нарушается условие (C2). Это нарушение условий выводимости проявляется в любопытном порочном круге при попытке доказательства непротиворечивости системы в самой системе. Пусть  $T \propto Con^*$ . Вопрос заключается в том, является ли это утверждение по-настоящему информативным о непротиворечивости  $T$ . Пусть  $T$  – непротиворечива, в этом случае  $Prf^*$  выражает отношение доказуемости, и тогда  $Con^*$  может быть интерпретировано как выражение непротиворечивости  $T$ . Но если  $T$  противоречиво, тогда  $Prf^*$  не выражает отношения доказуемости, и  $Con^*$  не может быть интерпретировано как выражение непротиворечивости.

Таким образом, оказывается, что мы уже должны знать, непротиворечива ли  $T$ , перед тем, как интерпретировать вывод  $Con^*$ . Но это означает, что сам по себе вывод  $Con^*$  не может нам сказать, действительно ли непротиворечива  $T$ . Графически эту ситуацию можно изобразить так (приведенное ниже является не формулой, а эвристической диаграммой):

$$(T \propto Con^*) \Rightarrow T \text{ непротиворечива.}$$

Чтобы получить заключение « $T$  непротиворечива», мы должны заранее знать, что  $T$  в « $T \propto Con^*$ » уже непротиворечива, так что искомое заключение уже предполагается всылке. Мы имеем тут если не порочный круг, то тавтологию. Так что доказательство непротиворечивости с  $Con^*$  не может ничего сказать о действительной непротиворечивости  $T$ .

Это обстоятельство очень важно для опровержения распространенной интерпретации Второй теоремы, согласно которой невозможно доказать непротиворечивость формальной системы в рамках самой этой системы. Крайзель приводит примеры формальных систем, в которых доказуема их собственная непротиворечивость, добавляя следующее замечание:

Кстати, с философской точки зрения интересно, что хотя в учебниках логики редко рассматриваются [такие] системы [которые] хорошо, хотя и грубовато, отражают один существенный метод, используемый на практике для проверки доказательств, а именно, сравнение с уже накопленными знаниями... [3. С. 47].

Это означает, что поиски  $Con^*$  имеют мотивацию в реальном математическом исследовании. Однако доказательства подобного рода не являются «каноническими» в следующем смысле. Что включает в себя такое гипотетическое непротиворечивое доказательство предложения  $\phi$  в аксиоматизированной системе  $T$ ? Технически это означает, что имеется вывод с геделевым номером  $m$  в системе  $T$  предложения  $\phi$  и в то же время не существует вывода в системе  $T$  ложного предложения такого, чтобы геделев номер такого вывода был меньше  $m$ . Но именно это обстоятельство отражено в структуре предиката доказуемости  $Prov^*$ . Изначально предполагается, что  $\Box_T \phi$  или же

$\text{Prov}_T(\ulcorner \varphi \urcorner)$  удовлетворяет условиям выводимости Гильберта – Бернайса. Выражение  $\text{Prov}_T(\ulcorner \varphi \urcorner)$  определено в терминах предикатного выражения  $\text{Prf}_T$ , которое канонически схватывает неформализованный предикат  $\text{Prf}_T$ . Это означает, что  $\text{Prf}_T(m, n)$ , где  $m$  есть геделево число доказательства утверждения с геделевым числом  $n$ , отражает детали именно такого соотношения  $m$  и  $n$ , и детали такого соотношения должны соблюдаться намеренно. В этом смысле  $\text{Prf}_T$  имеет намеренное значение, и в этом смысле Вторая теорема является интенциональной [19. Р. 263].

Отличие интенциональной трактовки от экстенциональной состоит в различии  $\text{Prov}_T$  и  $\text{Prov}^*_T$ . Формальное выражение отличия состоит в том, что  $\text{Prov}^*_T$  не удовлетворяет условиям выводимости Гильберта – Бернайса. Нарушение этих условий обесценивает формальную возможность доказательства непротиворечивости теории  $T$  в рамках самой  $T$ . Действительно, для обоснованной системы, т.е. когда все доказуемые утверждения являются истинными, выводимое в  $T$  гипотетическое  $\text{Con}^*$  является также истинным утверждением для любой теории  $T$ . Но именно то обстоятельство, что на  $T$  не накладывается никаких ограничений, обесценивает  $\text{Con}^*$ . В самом деле, для противоречивой системы  $T$  существует гипотетическое доказательство непротиворечивости  $\varphi$ ,  $\varphi \rightarrow \psi$  и  $\neg \psi$ . Это происходит потому, что гипотетическое выведение в  $T$  утверждения  $\psi$  – например, построения в цепочке доказательства  $\varphi$  после доказательства  $\varphi \rightarrow \psi$  и использования *modus ponens* – может иметь больший геделев номер, чем доказательство  $\neg \psi$ . При этом, в частности, нарушается условие (C2):  $T \sqcup \sqcup (\varphi \rightarrow \psi) \rightarrow (\sqcup \varphi \rightarrow \sqcup \psi)$ . Таким образом, истинность  $\text{Con}^*$  открыта возможности того, что теоремы, доказанные при гипотетическом доказательстве их непротиворечивости, семантически противоречивы. Это означает, что не существует интерпретации логического аппарата, которая делала бы такие теоремы истинными. В качестве примера Смит заключает, что, если для некоторой обоснованной теории  $T$ ,  $T \sqcup \text{Con}^*_{\text{ZFC}}$ , что делает тем самым  $\text{Con}^*_{\text{ZFC}}$  истинным, это вовсе не означает, что теория  $\text{ZFC}$  семантически непротиворечива, и даже синтаксически непротиворечива в обыденном смысле. Таким образом, то, из какого рода предикатов доказуемости конструируются предложения о непротиворечивости, является важнейшим обстоятельством, определяющим доказуемость или недоказуемость предложений непротиворечивости.

Здесь можно подвести итоги сопоставления Первой и Второй теорем Геделя о неполноте. Чисто формальное различие между ними сводится к следующим двум пунктам [19. Р. 265]:

В Первой теореме мы можем использовать, наряду с канонической правильно построенной формулой  $\text{Prf}_T$ , ппф  $\text{Prf}^*_T$ ; обе представляют, или схватывают, неформальный предикат доказуемости  $\text{Prf}_T$ . Далее, образуем соответствующий предикат доказуемости  $\text{Prov}^*_T$ . Рассмотрим неподвижную точку  $\alpha$  для предиката  $\neg \text{Prov}^*_T(x)$ . Тогда можно показать, что для достаточно богатой непротиворечивой системы  $T$ ,  $T \neg \sqcup \alpha$ , и для  $\omega$ -непротиворечивой системы  $T$ ,  $T \neg \sqcup \neg \alpha$ . Самым важным в данном случае является то, что доказательство этого факта зависит только от того факта, что  $\text{Prf}^*_T$  схватывает  $\text{Prf}_T$ , но никак не зависит от того, как это делается.

Во Второй теореме, для доказательства, что  $T \rightarrow \Box \text{Con}_T$  (в предположении  $\text{nice } T$ ), дело обстоит тоньше. Нам нужно начинать с ппф  $\text{Prf}_T$ , которая имеет «правильное значение», т.е. канонически представляет  $\text{Prf}_T$ , и затем уже из него формировать каноническое предложение  $\text{Con}_T$ . Если мы представляем  $\text{Prf}_T$  неверно, используя  $\text{Prf}^*_T$ , тогда соответствующее предложение  $\text{Con}^*T$  может быть доказуемо в  $T$ . Причины, по которым такая доказуемость нежелательна, описаны выше.

Упоминание выше о непротиворечивости в обыденном смысле также является ключевым, поскольку в значительной степени интенциональность Второй теоремы есть вопрос о переводе формальных результатов на обыденный язык, что, собственно, и придает теоремам Геделя их значимость в программах философии математики. Когда речь заходит об интерпретации математических результатов, всегда имеется в виду некоторого рода перевод в постижимые идиомы соответствующей семантики, коль скоро речь идет о самореференции. Контраст между двумя видами утверждений существен. Ауэрбах [9. Р. 337–338] сопоставляет математическую формулировку Первой теоремы Геделя о неполноте

(1) Не существует непротиворечивого полного аксиматизируемого расширения  $Q$  с обычным объяснением сути теоремы

(2) Любая достаточно сильная непротиворечивая формальная система арифметики неполна.

В то время как (1) является математически доказуемым утверждением, (2) является пересказом идей теоремы. Исходя из контраста между ними, неизбежно возникает вопрос, каковы основания верить в (2) как в выражение Первой теоремы Геделя. Это не значит, что (2) является произвольным пересказом технического результата. Существуют утверждения, являющиеся частью математического дискурса, но не являющиеся «чисто» математическими, например Тезис Черча, формулировка которого не претендует на точность и полную ясность. Эта аналогия может быть перенесена на обсуждение Второй теоремы Геделя о неполноте. Ее содержание передается аналогом предложения (2), т.е. содержательно, следующим образом:

(3) Если  $Q$  достаточно сильная непротиворечивая формальная система арифметики, тогда любое предложение  $Q$ , утверждающее, что  $Q$  непротиворечиво, не выводимо в  $Q$ .

Обращаясь к соотношению (1) и (2) как образцу соотношения точного математического результата и его философской интерпретации и учитывая, что (3) является интерпретацией, представляет интерес, каков аналог (1) для Второй теоремы Геделя. Дело в том, что в отличие от Первой теоремы Геделя, математическая формулировка которой вполне постижима и без ее «пересказа», Вторая теорема постижима скорее в «пересказе», так что есть определенная асимметрия между двумя теоремами в этом отношении. Это и является еще одним источником интенциональности Второй теоремы.

### Литература

1. Гильберт Д., Бернайс П. Основания математики. Т. 2: Теория доказательств. М.: Наука, 1979.
2. Smoryński C. Review: Peter Smith. An Introduction to Gödel's Theorem. Cambridge: Cambridge University Press, 2007 // *Philosophia Mathematica*. 2010. Vol. 18, № 1. P. 122–127.

3. Крайзель Г. Биография Курта Геделя. М.: Институт компьютерных исследований, 2003.
4. Feferman S. Arithmetization of Metamathematics in a General Setting // *Fundamenta Mathematicae*. 1960. Vol. 49. P. 35–92.
5. Smorynski C. The Development of the Self-Reference: Lob's Theorem // *Perspectives on the History of Mathematical Logic* / ed. T. Drucker. Berlin: Birkhauser, 1991. P. 110–133.
6. Castonguay C. Meaning and Existence in Mathematics. Springer, 1972.
7. Curtis F. The Autonomy of Mathematical Knowledge: Hilbert's Program Revisited, Cambridge: Cambridge University Press, 2009.
8. Detlefsen M. On Interpreting Gödel's Second Incompleteness Theorem // *Journal of Philosophical Logic*. 1979. Vol. 8. P. 297–315.
9. Auerbach D. Intensionality and the Gödel Theorem // *Philosophical Studies*. 1985. Vol. 48. P. 337–351.
10. Kreisel G. On the Problem of Henkin // *Proceedings of Netherland Academy of Science*. 1953. Vol. 56. P. 405–406.
11. Halbach V., Visser A. Self-Reference in Arithmetic I // *The Review of Symbolic Logic*. 2014. Vol. 7, №. 4. P. 671–691.
12. George A., Velleman D. *Philosophies of Mathematics*. Oxford: Blackwell, 2002.
13. Buldt B. On Fixed Points, Diagonalization, and Self-Reference // *Von Rang and Namen. Essays in Honour of Wolfgang Spohn* / eds. Freitag W. et al. Munster: Mentis, 2016.
14. Sereny G. How do We Know that the Godel Sentence of a Consistent Theory is True? // *Philosophia Mathematica* (III). 2001. Vol. 19. P. 47–73.
15. Kreisel G. *Mathematical Logic* // *Lectures on Modern Mathematics 111* / ed. T.L. Saaty. N.Y.: Wiley, 1965.
16. Henkin L. Problem // *Journal of Symbolic Logic*. 1952. Vol. 17.
17. Kreisel G. Gödel's Excursion into Intuitionistic Logic // *Gödel Remembered* / ed. P. Weingartner, L. Schmetterer. Napoli: Bibliopolis, 1987. P. 65–179.
18. Gödel K. Diskussion zur Grundlegung der Mathematik // *Erkenntnis*. Bd. 2. 1931–32. S. 147–151.
19. Smith P. *An Introduction to Gödel's Theorems*. Second Edition. Cambridge: Cambridge University Press, 2013.
20. Mostowski A. *Sentences Undecidable in Formalized Arithmetic: An Exposition of Theory of Kurt Gödel*. Amsterdam: North-Holland Publishing Co, 1952.

**Vitaly V. Tselishchev.** Novosibirsk State University; Institute of Philosophy and Law, Siberian Division, Russian Academy of Science (Novosibirsk, Russian Federation)

E-mail: leitval@gmail.com

DOI: 10.17223/1998863X/40/10

# **INTENSIONALITY OF THE GÖDEL'S SECOND INCOMPLETENESS THEOREM**

**Key words:** intensionality, Gödel's theorem, self-reference, consistency, proof

The difference between the First and Second Gödel's theorem is discussed in perspective of possible rehabilitation of the Hilbert Program. The essence of the difference lies in the fact that if the First theorem is extensional, then the Second theorem is intensional. The formalization of the consistency concept used in the proof of the Second Theorem is reconsidered, particularly, the other ways of expressing the consistency property that are not subject to this theorem. In this paper, we investigate what can be considered a formal expression of a meaningful concept of consistency. The formalization of the consistency concept, used in the proof of the Second Theorem, is contested with the presentation of other ways of expressing the consistency property, for which the second theorem does not hold. It is investigated what can be considered a formal expression of the concept of consistency.

## **References**

1. Gilbert, D. & Bernays, P. (1979) *Osnovaniya matematiki* [Fundamentals of mathematics]. Translated from German by N.M. Nagorny. Vol. 2. Moscow: Nauka.
2. Smorynski, C. (2010) Review: Peter Smith. *An Introduction to Gödel's Theorem*. Cambridge: Cambridge University Press, 2007. *Philosophia Mathematica*. 18(1). pp. 122–127.
3. Kreisel, G. (2003) *Biografiya Kurta Gedelya* [Kurt Gödel's Biography]. Translated from German by G.E. Mints, D.P. Skvortsov, E. Z. Skvortsova. Moscow: Institute for Computer Research.

4. Feferman, S. (1960) Arithmetization of Metamathematics in a General Setting. *Fundamenta Mathematicae*. 49. pp. 35–92. DOI: 10.2307/2269834
5. Smoryński, C. (1991) The Development of the Self-Reference: Lob's Theorem. In: Drucker, T. (ed.) *Perspectives on the History of Mathematical Logic*. Berlin: Birkhauser. pp. 110–133.
6. Castonguay, C. (1972) *Meaning and Existence in Mathematics*. Springer.
7. Curtis, F. (2009) *The Autonomy of Mathematical Knowledge: Hilbert's Program Revisited*. Cambridge: Cambridge University Press.
8. Detlefsen, M. (1979) On Interpreting Gödel's Second Incompleteness Theorem. *Journal of Philosophical Logic*. 8. pp. 297–315. DOI: <https://doi.org/10.1007/BF00258433>
9. Auerbach, D. (1985) Intensionality and the Gödel Theorem. *Philosophical Studies*. 48. pp. 337–351.
10. Kreisel, G. (1953) On the Problem of Henkin. *Proceedings of Netherland Academy of Science*. 56. pp. 405–406. DOI: <https://doi.org/10.2307/2268627>
11. Halbach, V. & Visser, A. (2014) Self-Reference in Arithmetic I. *The Review of Symbolic Logic*. 7(4). pp. 671–691. DOI: 10.1017/S1755020314000288
12. George, A. & Velleman, D. (2002) *Philosophies of Mathematics*. Oxford: Blackwell.
13. Buldt, B. (2016) On Fixed Points, Diagonalization, and Self-Reference. In: Freitag, W. et al. (eds) *Von Rang and Namen. Essays in Honour of Wolfgang Spohn*. Munster: Mentis.
14. Sereny, G. (2001) How do We Know that the Godel Sentence of a Consistent Theory is True? *Philosophia Mathematica (III)*. 19. pp. 47–73. DOI: 10.1093/phimat/nkq028
15. Kreisel, G. (1965) Mathematical Logic. In: Saaty, T.L. (ed.) *Lectures on Modern Mathematics III*. New York: Wiley.
16. Henkin, L. (1952) Problem. *Journal of Symbolic Logic*. 17.
17. Kreisel, G. (1987) Gödel's Excursion into Intuitionistic Logic. In: Weingartner, P. & Schmetterer, L. (eds) *Gödel Remembered*. Napoli: Bibliopolis. pp. 65–179.
18. Gödel, K. (1931–1932) Diskussion zur Grundlegung der Mathematik [Discussion on the foundation of mathematics]. *Erkenntnis*. 2. pp. 147–151.
19. Smith, P. (2013) *An Introduction to Gödel's Theorems*. 2nd ed. Cambridge: Cambridge University Press.
20. Mostowski, A. (1952) *Sentences Undecidable in Formalized Arithmetic: An Exposition of Theory of Kurt Gödel*. Amsterdam: North-Holland Publishing Co.