2010 г. сентябрь — октябрь *т. 65, вып. 5 (395)*
УСПЕХИ МАТЕМАТИЧЕСКИХ НАУК

УДК 510.2+510.6

Теоремы Гёделя о неполноте и границы их применимости. I

Л. Д. Беклемишев

Дан обзор результатов, связанных с теоремами Гёделя о неполноте и границами их применимости. В первой части обсуждаются формулировки самого Гёделя, а также современные усиления первой теоремы о неполноте. Сравниваются между собой различные формы и доказательства этой теоремы. Рассматриваются результаты о неполноте, связанные с алгоритмическими проблемами, и обсуждаются математически естественные примеры недоказуемых утверждений.

Библиография: 68 названий.

Ключевые слова: теоремы Гёделя, неполнота, доказательство, вычислимость.

Содержание

Работа выполнена при поддержке РФФИ и программы “Ведущие научные школы”.

1. Введение

Теоремы Гёделя о неполноте являются общепризнанным достижением математической мысли XX века, заложившим фундамент математической логики. Методы, созданные Гёделем, были одним из решающих факторов, приведших к точному математическому определению алгоритма и, в конечном итоге, к созданию компьютеров.

В то же время, по самой своей природе теоремы Гёделя затрагивают вопросы, далеко выходящие за пределы собственно математики и связанные с такими волнующими воображение темами, как загадка природы человеческого разума, проблема познания или искусственный интеллект. В этом своем качестве теоремы Гёделя – или скорее их интерпретации – сыграли заметную роль в формировании общего интеллектуального контекста XX века. В результате теоремы Гёделя стали одним из самых известных за пределами самой математики математических открытий минувшего столетия.[1]

Теоремам Гёделя о неполноте посвящена обширнейшая литература, от сугубо специальной до педагогической, научно-популярной и художественной. К последней категории, в частности, относятся такие мировые бестселлеры, как “Gödel, Escher, Bach: the Eternal Golden Braid” Дугласа Хофштадтера или “The Emperor’s New Mind” и “Shadows of the Mind” Роджера Пенроуза. В море информации, посвященной теоремам Гёделя, преобладают доступные для достаточно широкого круга читателей изложения этих теорем (прежде всего, его более простой первой теоремы). В некотором смысле, сама природа этих результатов позволяет бесконечно варьировать и приспосабливать идеи Гёделя ко вкусам того или иного автора.

В данном обзоре мы не ставим себе целью дать еще одно доступное изложение теорем Гёделя. Нам хотелось бы восполнить пробел другого рода: познакомить читателя с современными обобщениями теорем Гёделя о неполноте, прежде всего его второй теоремы, и с разнообразными тонкими математическими вопросами вокруг этих результатов.

Поскольку теоремы Гёделя относятся к классике математической логики, мы не приводим всех деталей доказательств, которые в той или иной форме достаточно хорошо известны. Вместо этого значительное внимание уделяется обсуждению и поискам оптимальных формулировок результатов Гёделя, а также сравнению различных доказательств между собой. Мы изложим результаты, связанные с вопросом о границах применимости теорем Гёделя о неполноте, накопленные в математической логике, в том числе и за самое последнее время.

Обзор естественным образом разбит на две части, посвященные соответственно первой и второй теоремам Гёделя.

Подходящий контекст для первой теоремы Гёделя был создан развитием теории вычислимых функций, и в рамках этой теории ее роль стала достаточно понятной. Однако вторая теорема Гёделя в полной мере в этот контекст не

[1]Косвенным свидетельством этого может служить тот факт, что Курт Гёдель был признан – справедливо или нет – одной из ста наиболее влиятельных личностей XX в. по версии журнала Time.

укладывается и в целом более проблематична. На начальном этапе преобладала точка зрения, согласно которой вторая теорема Гёделя является некоторым “довеском” к первой теореме, всего лишь указывающим явную форму того самого независимого утверждения, о существовании которого говорит первая теорема. До некоторой степени эта точка зрения распространена и сейчас.

В то же время, после работ Крайзеля, Лёба, Фефермана, де Йонга, Сморинского и других исследователей, изучавших обобщения второй теоремы Гёделя, стало понятно, что природа этих двух утверждений существенно разная. Вторая теорема Гёделя оказывается в большей мере связанной с модально-логическими свойствами формулы, выражающей доказуемость, и с эффектом самоописания в арифметике.

С нашей точки зрения, еще рано выносить окончательное суждение о том, что представляет собой “правильный” контекст для второй теоремы Гёделя о неполноте. В то же время, в этой области накоплено много частных результатов, проливающих свет на ее роль и особенности. Во второй части настоящего обзора мы предполагаем познакомить читателя с этими результатами и остающимися здесь открытыми вопросами.

В первой части мы обсуждаем первую теорему Гёделя и ее различные формы. Этот материал совершенно традиционный, однако мы решили включить его достаточно подробный обзор по двум причинам. Во-первых, проблемы, связанные со второй теоремой Гёделя, можно оценить в полной мере, лишь имея очень ясное представление о различных аспектах первой теоремы. В то же время, многие существующие изложения этого материала часто бывают слишком однобокими. Во-вторых, в литературе можно найти много разных утверждений, называемых теоремой Гёделя, не говоря уже об их различных доказательствах. Нам хотелось бы отчасти привести в систему этот материал, рассматривая его с некоторой единой позиции.

Мне хотелось бы поблагодарить Владимира Андреевича Успенского и Альберта Виссера за обсуждение некоторых затрагиваемых здесь вопросов и Сергея Ивановича Адяна за внимательное чтение и критику рукописи. Их работы также во многом повлияли на взгляды автора, нашедшие отражение в этом обзоре.

2. Первая и вторая теоремы о неполноте по Гёделю

В основополагающей работе [1] Гёдель доказал свои теоремы для определенной формальной системы P, родственной *Principia Mathematica* Рассела–Уайтхеда и основанной на простой теории типов над натуральным рядом и аксиомах Дедекинда–Пеано. В то время *Principia Mathematica* была, пожалуй, самой известной из систем аксиом, предназначенных для формализации математики. Гёдель также объяснил, что его результат распространяется и на другие аксиоматические системы, включающие теорию множеств Цермело–Френкеля, теорию множеств фон Неймана и другие теории, “разрабатываемые в последнее время Д. Гильбертом и его учениками”. Мы не будем давать точное определение системы P, формулировка которой несколько громоздка. Более сильные результаты будут получены для стандартного в наше время языка арифметики первого порядка.

Простейшая формулировка первой теоремы Гёделя о неполноте говорит о том, что существует предложение, не доказуемое и не опровержимое в рамках рассматриваемой теории P. Вторая теорема Гёделя утверждает, что в качестве такого предложения можно взять формализацию в P утверждения о ее собственной непротиворечивости.

Неполнота таких теорий, как P, или теорий множеств, специально созданных для аксиоматизации всей математики, находилась в резком противоречии с преобладавшими в то время взглядами. Вторая же теорема Гёделя ставила под сомнение возможность реализации наиболее важного положения так называемой *программы Гильберта* (см. [2]), декларированной целью которой было установление непротиворечивости математики (анализа и теории множеств) финитными средствами. Эта задача рассматривалась представителями школы Гильберта как центральная проблема математической логики. В самом деле, из второй теоремы Гёделя вытекает, что "финитные средства", способные установить непротиворечивость математики, невозможно формализовать даже в рамках очень сильной системы P.[2]

В принципе, неполнота какой-либо конкретной теории, например P, может означать всего лишь, что "не были учтены" какие-то нужные аксиомы. (Например, так обстояло долгое время дело с аксиоматизацией элементарной геометрии.) Чтобы показать, что в данном случае мы имеем дело с существенно другой и более драматичной ситуацией, Гёдель формулирует свою первую теорему в более общей форме, которая говорит о *принципиальной непополняемости* системы P.

По смыслу достаточно близко к оригиналу *первая теорема Гёделя о неполноте* [1; теорема VI] может быть сформулирована следующим образом.

Теорема 1. *Пусть T – формальная теория, удовлетворяющая следующим условиям:*

(i) *T сформулирована в языке P;*
(ii) *T получается добавлением к системе P примитивно рекурсивного множества аксиом;*
(iii) *T является ω-непротиворечивой.*

Тогда T неполна, т. е. существует предложение φ, для которого ни φ, ни $\neg\varphi$ не доказуемы в T.

Поясним участвующие здесь понятия и сами условия (i)–(iii). Следует сразу отметить, что ни одно из этих условий с современной точки зрения не является

[2]Гёдель выражает это обстоятельство любопытным образом, видимо, не желая вступать в философскую полемику: "*Следует особенно подчеркнуть, что теорема* IX [*вторая теорема о неполноте*] *не находится в противоречии с гильбертовской формалистской точкой зрения. Ведь* [...] *можно себе представить, что существуют финитные доказательства, которые* нельзя *формализовать в P*". В настоящее время принято считать, что финитные доказательства формализуемы даже в гораздо более слабой арифметике Пеано первого порядка. В то же время, существуют и альтернативные точки зрения по этому вопросу, начало которым положил сам Гильберт (см. [2]).

ни наиболее общим, ни наиболее математически естественным. (Причины этого коротко обсуждаются ниже, см. также соответствующие замечания в [3] и [4].) Однако эти условия интересны, они отражают первоначальное понимание Гёделем своего открытия в 1931 г., и к тому же их использование в значительной степени закреплено в последующей литературе.

В условии (ii) участвует понятие примитивно рекурсивной функции.

Определение 1. Функция $f\colon \mathbb{N}^k \to \mathbb{N}$ называется *примитивно рекурсивной*, если она может быть получена из константы 0, функции следования $S(x) = x+1$ и проектирующих функций $I_n^m(x_1,\dots,x_n) = x_m$ с помощью операций композиции (подстановки) и примитивной рекурсии. Функция f получена из g и h примитивной рекурсией, если

$$\begin{cases} f(0,\vec{x}\,) = g(\vec{x}\,), \\ f(n+1,\vec{x}\,) = h\bigl(f(n,\vec{x}\,),n,\vec{x}\,\bigr). \end{cases}$$

Множество $R \subseteq \mathbb{N}^k$ примитивно рекурсивно, если утверждение $\vec{x} \in R$ равносильно $f(\vec{x}\,) = 0$ для некоторой примитивно рекурсивной функции f. (Как это принято в логике, утверждение $\vec{x} \in R$ мы будем также записывать $R(\vec{x}\,)$.)

Условие (ii) предполагает, что заранее фиксировано некоторое взаимно однозначное кодирование натуральными числами (гёделева нумерация) объектов языка теории P, таких как переменные, термы, формулы и т. д. Гёдель показал, что при некотором естественном выборе такого кодирования отношение "x есть код вывода формулы с кодом y в P" будет примитивно рекурсивным.

Условие (ii) говорит о том, что множество кодов аксиом T при выбранном кодировании должно быть примитивно рекурсивным. В силу того, что T строится на базе P, а дедуктивный механизм P примитивно рекурсивен, отсюда также следует примитивная рекурсивность отношения "x есть код вывода формулы y в T". Требование (ii) является достаточно общим и для всех формальных теорий, рассматриваемых на практике, выполняется. Гёдель отметил, что этому условию заведомо удовлетворяют теории, задаваемые конечным числом аксиом или схем аксиом. Теории, удовлетворяющие условию (ii), сейчас принято называть *примитивно рекурсивно аксиоматизированными.*

Определение 2. Теория T называется *ω-непротиворечивой*, если не существует такой формулы $\varphi(x)$ (переменная x пробегает натуральные числа), что одновременно выполнены условия

(i) $T \vdash \exists x\, \varphi(x)$;

(ii) $T \vdash \neg\varphi(\underline{0}), \neg\varphi(\underline{1}), \dots$.

Напомним, что для теорий T в языке P натуральное число n изображается замкнутым термом $S(S(\dots S(0)\dots))$ (n раз), который мы обозначаем $\underline{n}$ и называем *нумералом.* Последовательность нумералов $(\underline{n}_1,\dots,\underline{n}_k)$ сокращенно записываем как $\vec{\underline{n}}$.

Условие ω-непротиворечивости усиливает требование непротиворечивости теории T. В свою очередь, оно следует из предположения о корректности T, т. е. об истинности всех теорем T в модели с носителем $\mathbb{N}$. Однако сам Гёдель не рассматривает семантическое понятие корректности в своей работе.

Условие (iii) также не является существенно ограничительным: теории, которые ему *не* удовлетворяют, можно рассматривать как близкие к противоречивым, т. е. как патологический случай. Фактически, Гёдель говорит, что при выполнении естественных и широких требований (i) и (ii) мы имеем выбор из двух "неприятных" возможностей: либо теория T неполна, либо ω-противоречива.

Что касается условия (i), то оно является достаточно ограничительным. Далеко не все естественные теории, такие, например, как теория множеств, формулируются буквально в языке P. В связи с этим в комментариях к теореме VI Гёдель отмечает, что в условиях теоремы достаточно вместо (i) потребовать, чтобы в теории T были *разрешимы* (entscheidungsdefinit)[3] все примитивно рекурсивные отношения, а вместо (ii) потребовать разрешимости в T множества кодов аксиом теории T.

Определение 3. Отношение R называется *разрешимым в* T, если существует формула $\varphi_R(\vec{x}\,)$ такая, что для всех наборов чисел $\vec{n}$

$$R(\vec{n}\,) \Rightarrow T \vdash \varphi_R(\underline{\vec{n}}\,), \tag{1}$$

$$\neg R(\vec{n}\,) \Rightarrow T \vdash \neg\varphi_R(\underline{\vec{n}}\,). \tag{2}$$

Центральным элементом доказательства Гёделя была теорема о разрешимости в P всех примитивно рекурсивных отношений (которую он, однако, обосновывает лишь схематически).[4] В частности, это позволяет ему выразить независимое утверждение для теории T в виде $\forall x\ \varphi_R(x)$, где R примитивно рекурсивно. Дополнительная идея кодирования конечных последовательностей чисел с помощью так называемой *β-функции Гёделя*, основанной на китайской теореме об остатках, показывает [1; теорема VII], что всякое примитивно рекурсивное отношение эквивалентно (в P) некоторой арифметической формуле, т. е. формуле языка первого порядка в сигнатуре с константой 0, функциями S, $+$ и $\cdot$ и отношением равенства. Тем самым Гёдель продемонстрировал, что построенное им независимое предложение для теории T, каким бы ни был язык самой теории T, относится к элементарной арифметике.

Обсуждение. Для того чтобы правильно оценить условия теоремы 1 с современной точки зрения, необходимо учитывать следующие обстоятельства.

Во-первых, теория вычислимости, которая дает правильный контекст для первой теоремы Гёделя, в то время еще не была создана. Фактически, именно теоремы Гёделя в большой степени способствовали созданию этой теории: в лекциях Гёделя 1934 г. [5] введено понятие общерекурсивной функции, эквивалентность которого с понятием (всюду определенной) вычислимой функции –

[3]Это понятие разные источники называют также *бинумеруемостью*, *представимостью* или *определимостью в теории*.

[4]Строгое обоснование этой теоремы приведено в лекциях 1934 г. для некоторой теории второго порядка, родственной системе P [5]. Интересно отметить, что благодаря наличию переменных по множествам в языке P эту теорему удается доказать существенно проще, чем для принятого в настоящее время языка арифметики первого порядка.

в других формулировках – была впоследствии установлена Чёрчем, Тьюрингом и Клини. Сам Гёдель отметил в добавлении к английскому переводу 1963 г. своей работы важную роль А. М. Тьюринга в том, что “теперь можно дать вполне общую версию теорем VI и IX” (т. е. первой и второй теоремы Гёделя, см. [6; с. 195]).

Во-вторых, Гёдель избегает использования семантических понятий, в частности, понятия модели, понятий истинности и выразимости в модели. Это связано, по-видимому, с двумя обстоятельствами. Прежде всего, до работы Гёделя полной ясности относительно различий между понятиями доказуемости и истинности для таких теорий, как P, не было. Более того, семантические понятия в логике в целом были под подозрением из-за всем известных парадоксов. Важная работа А. Тарского [7], касающаяся анализа понятия истинности для формализованных языков и в значительной мере снявшая такие подозрения, была опубликована лишь в 1933 г. (а на немецком языке в 1935 г.) Этим можно объяснить и возникновение условия ω-непротиворечивости вместо более фундаментального условия семантической корректности. Также не было строгого понятия интерпретации одной теории в другой (сформировавшегося значительно позже под влиянием работ Тарского), что привело, в частности, к ограничению в формулировке теорем о неполноте теориями в языке P.

С другой стороны, работа Гёделя была написана, прежде всего, в расчете на читателей, разделяющих взгляды Гильберта на основания математики. Это также повлияло на ее стиль и выбор некоторых формулировок в ущерб более естественным семантическим. Гёдель стремился показать, что теоремы о неполноте, сами по себе, относятся к финитной математике и не связаны с чем-то иррациональным, с “неправомерным использованием бесконечности” и тому подобным.

В-третьих, в работе Гёделя также еще нет четкого разделения понятий теории и метатеории. Хотя термины “метаязык” и “метатеория” восходят к Гильберту, принципиальность этих понятий для рассматриваемого круга вопросов стала ясна благодаря работе Гёделя и уже упомянутой работе Тарского.

По-видимому, корректное прочтение работы Гёделя состоит в неявном выборе самой системы P в качестве метатеории, в которой формализуемы все содержательные утверждения работы. На это указывает особый формализованный стиль формулировок всех утверждений статьи (с использованием особого шрифта для формальных аналогов содержательных синтаксических понятий), начиная с теоремы V о разрешимости примитивно рекурсивных отношений в P.[5] Сам Гёдель указывает на формализуемость в P значительной части утверждений статьи, описывая схему доказательства своей второй теоремы. Возможно, выбор столь формального стиля изложения был продиктован именно желанием помочь читателю поверить в возможность такой формализации.

[5] В своем глубоком комментарии [4] к работе Гёделя, Клини также отмечает это обстоятельство, но не объясняет его, называя просто “пристрастием (propensity) Гёделя говорить в терминах своих номеров”.

В-четвертых, понятие разрешимости по Гёделю играет роль подходящей замены семантическому понятию определимости. Он подчеркивает, что это понятие не апеллирует к "содержательному значению (inhaltliche Deutung) формул системы P". Однако все же можно заметить, что это понятие неявно апеллирует к содержательной интерпретации примитивно рекурсивных схем, поскольку формула φ_R фактически строится по примитивно рекурсивной схеме, определяющей R, а не по содержательному отношению, выражаемому R. Стоит отметить, что Гёдель иногда не вполне корректно отождествляет примитивно рекурсивные функции и схемы, например при определении *уровня* примитивно рекурсивной функции (по этому поводу см. также комментарий Клини). С философской точки зрения, семантическая интерпретация примитивно рекурсивных функций в каком-то смысле яснее и проще, чем таковая для произвольных формул языка P (высшего порядка).

В связи с понятием разрешимости в теории необходимо отметить еще одно важное обстоятельство. Позднейший анализ этого понятия (см. [8]) показал, что для непротиворечивых, примитивно рекурсивно аксиоматизированных теорий, содержащих P (или даже намного более слабую арифметику Робинсона Q), это понятие равнообъемно с понятием алгоритмической разрешимости. Тем самым, более общая формулировка первой теоремы Гёделя (данная им как комментарий на с. 190 работы [1][6]), по существу, относится к тому же классу теорий, что и современная формулировка, рассматриваемая в следующем разделе. Приведем формулировку этого комментария здесь более явно, чем у самого Гёделя, для сравнения с последующими утверждениями.

Утверждение 1 (комментарий Гёделя). *Пусть T – формальная теория, удовлетворяющая следующим условиям*:

(i) *в T разрешимы все примитивно рекурсивные отношения*;

(ii) *множества аксиом и правил вывода системы T (т. е. отношение непосредственного следования в T) примитивно рекурсивны или хотя бы разрешимы в T*;

(iii) *T является ω-непротиворечивой.*

Тогда T неполна.

Здесь Гёдель не уточняет, о теориях в каком языке идет речь, но все три условия – в том числе и (i) – формулируются в терминах абстрактного отношения "непосредственного следования" и соответствующего ему понятия выводимости. Отсюда очень недалеко до абстрактного понятия формальной системы как алгоритма, порождающего теоремы T исходя из аксиом по правилам вывода, множество которых примитивно рекурсивно (или хотя бы разрешимо). Поэтому выбор языка в данном случае не играет большой роли, лишь бы для

[6]Предшествующее замечание Гёделя на с. 189 обобщает условие примитивно рекурсивной аксиоматизируемости до разрешимости множества аксиом теории T в T.

него имело смысл понятие разрешимости в теории.[7] Разумеется, в явном виде этого понятия у Гёделя еще не было – вспомним ту роль, которую Гёдель отводил Тьюрингу и другим в последующем формировании этого понятия.

Отметим также, что утверждение 1 может относиться и к теориям, не являющимся перечислимыми. Например, в качестве T можно взять теорию первого порядка, аксиоматизируемую над P всеми формулами $\forall x\, \varphi(x)$ такими, что для каждого конкретного $n \in \mathbb{N}$ в P доказуемо $\varphi(\underline{n}\,)$ (однократное применение ω-правила Карнапа, см. [10]). Теории такого рода сам Гёдель в своей работе не рассматривал, тем не менее они играли и играют важную роль в исследованиях по основаниям математики.

3. Современные формулировки первой теоремы Гёделя

В результате возникновения общего понятия вычислимой функции в работах К. Гёделя, А. Чёрча, А. Тьюринга, Э. Поста, С. Клини стало ясно, что класс примитивно рекурсивных функций не является столь же фундаментальным, как класс произвольных вычислимых функций или класс перечислимых отношений. Соответственно, условие (ii) примитивно рекурсивной аксиоматизируемости теории T в первой теореме Гёделя естественно обобщить до условия *перечислимости* множества теорем теории T. Такое обобщение расширяет класс рассматриваемых аксиоматизаций, однако не влияет на класс рассматриваемых теорий, поскольку, как было позже замечено У. Крейгом [11], всякая перечислимая теория на базе логики первого порядка может быть задана примитивно рекурсивным множеством аксиом.

С. Клини [12] был, по-видимому, одним из первых, кто переосмыслил результаты Гёделя с точки зрения более общей теории вычислимости. Фактически, уже в этой ранней работе Клини содержатся основные идеи подхода к теоремам Гёделя на основе развиваемой им теории рекурсии, которые легли в основу практически всех дальнейших исследований вокруг теорем Гёделя. Там же Клини дал и общую теоретико-рекурсивную версию первой теоремы о неполноте (в семантическом варианте). С точки зрения эволюции формулировок, она занимает промежуточное положение между неформальными комментариями Гёделя к своим теоремам и более совершенными абстрактными формулировками, которые можно найти как у самого Клини в позднейших работах [13]–[15], так и у других авторов, например Р. Смальяна [16] и В. А. Успенского [17].

С другой стороны, Дж. Б. Россер [18] модифицировал конструкцию Гёделя таким образом, что установил неполноту теории T уже при условии ее обычной непротиворечивости. Эту более сильную и красивую форму первой теоремы Гёделя иногда называют теоремой Россера или теоремой Гёделя–Россера. (Конструкция самого Гёделя, тем не менее, сохранила свое значение в связи с доказательством второй теоремы о неполноте.)

[7] Для языков первого (или высшего) порядка, содержащих имена для натуральных чисел, понятие разрешимости, данное Гёделем, является вполне строгим. В общем случае можно сформулировать достаточно простые абстрактные требования к языку для того, чтобы понятие разрешимости в соответствующей формальной системе имело смысл. Это приводит, например, к понятию *представляющей системы* по Смальяну [9].

Дальнейшие технические улучшения первой теоремы о неполноте связаны с рядом результатов С. Клини, А. Мостовского, А. Тарского и Р. Робинсона. Базируясь, в частности, на работе Гёделя, Клини [13] и Мостовский [19] исследовали отношения, определимые в языке арифметики первого порядка, и ввели в рассмотрение иерархию арифметических классов Σ_n и Π_n.

А. Тарский инициировал систематическое исследование вопросов разрешимости для теорий первого порядка. В рамках этой программы в школе Тарского были получены весьма широкие обобщения результатов Гёделя. С одной стороны, были найдены очень слабые арифметические теории, такие как арифметика Робинсона Q, для которых была установлена существенная неразрешимость. Тем самым теорема Гёделя–Россера была распространена на более широкий класс теорий. С другой стороны, развитый Тарским метод интерпретаций позволил перенести эти результаты на другие языки, отличные от арифметического. Отметим, что одним из самых ранних и красивых результатов в этом направлении была теорема Куайна [20], установившая взаимную определимость элементарной арифметики натуральных чисел и теории конкатенации двоичных слов. В комбинации с методом интерпретаций теорема Гёделя–Россера стала одним из главных орудий доказательства алгоритмической неразрешимости самых разных теорий.

После выхода в начале 50-х годов фундаментальных монографий Клини [15] и Тарского, Мостовского и Робинсона [21] изложения теорем Гёделя о неполноте приобрели их стандартные формы, очень близкие к современным.

3.1. Общие формулировки. Сначала мы дадим очень общие формулировки синтаксических версий теорем Гёделя и Гёделя–Россера для абстрактных формальных систем. Эти формулировки имеют то преимущество, что применимы в большом числе ситуаций и не зависят от языка теории. Кроме того, они выявляют вычислительную суть первой теоремы Гёделя.

Начнем с определения абстрактной формальной системы.

Определение 4. *Формальной системой* называем тройку $S=(L,P,R)$, где $L\subseteq\mathbb{N}$ – разрешимое множество всех *предложений* системы, а $P,R\subseteq L$ – перечислимые множества *доказуемых* и *опровержимых* предложений соответственно. Считаем, что все три множества заданы фиксированными алгоритмами (машинами Тьюринга), обозначаемыми M_L, M_P и M_R.

Заметим, что мы отождествили формальные объекты системы и их гёделевы номера, поэтому соответствующие алгоритмы работают на натуральных числах. Система S называется *непротиворечивой*, если $P\cap R=\varnothing$, и *полной*, если она непротиворечива и $P\cup R=L$.

Заметим, что если формальная система S полна, то множества P и R разрешимы (теорема Поста). Поэтому для доказательства неполноты S достаточно установить алгоритмическую неразрешимость любого из этих двух множеств. Систему S называем *разрешимой*, если разрешимо P.

Неразрешимость системы S вытекает из существования перечислимого неразрешимого множества и естественного требования, чтобы S в каком-то смысле “выражала” все перечислимые множества. Весьма общее понятие выразимости для абстрактных формальных систем можно определить следующим образом.

Определение 5. Множество $A \subseteq \mathbb{N}$ *выразимо в* $S = (L, P, R)$, если существует всюду определенная вычислимая функция f такая, что

$$n \in A \iff f(n) \in P.$$

Система S называется *универсальной для перечислимых множеств*, если в ней выразимы все перечислимые множества.

Понятие выразимости для формальных систем есть не что иное, как известное в теории алгоритмов понятие m-сводимости к множеству доказуемых предложений S. Любое множество, выразимое в разрешимой системе, должно быть алгоритмически разрешимым. Таким образом, универсальные формальные системы неразрешимы и неполны. Этот факт можно считать некоторой абстрактной версией первой теоремы Гёделя о неполноте.

Теорема 2. *Если S – формальная система, универсальная для перечислимых множеств, то S неразрешима и неполна.*

В основе теоремы Гёделя–Россера лежит другой факт из теории вычислимых функций – существование неотделимой пары непересекающихся перечислимых множеств. Заметим, что с каждой упорядоченной парой непересекающихся множеств $A, B \subseteq \mathbb{N}$ можно связать частичную функцию $g\colon \mathbb{N} \to \{0, 1\}$, для которой

$$g(n) = \begin{cases} 0, & \text{если } n \in A, \\ 1, & \text{если } n \in B, \\ \text{не определено} & \text{в остальных случаях.} \end{cases}$$

Такое соответствие между парами и функциями взаимно однозначно. Оба множества A и B перечислимы, если и только если функция g вычислима.

Система $T = (L_T, P_T, R_T)$ называется *расширением* $S = (L_S, P_S, R_S)$, если $L_T \supseteq L_S$, $P_T \supseteq P_S$ и $R_T \supseteq R_S$. В этом случае функция g_T, соответствующая паре (P_T, R_T), является продолжением функции g_S.

Пару (A, B) называем (*рекурсивно*) *неотделимой*, если функция g не имеет всюду определенного вычислимого продолжения $g'\colon \mathbb{N} \to \{0, 1\}$. Формальная система $S = (L, P, R)$ *неотделима*, если такова пара (P, R).

В силу сказанного выше, неотделимые формальные системы неразрешимы и неполны, и таковыми являются любые их непротиворечивые расширения. Таким образом, неотделимость формальной системы является еще одним достаточным условием неполноты.

Определение 6. Пара (A, B) *отделима в системе* $S = (L, P, R)$, если существует вычислимая функция f такая, что

$$\begin{cases} n \in A \Rightarrow f(n) \in P, \\ n \in B \Rightarrow f(n) \in R. \end{cases}$$

Говорим, что S *отделяет пары перечислимых множеств*, если любая пара непересекающихся перечислимых множеств (A, B) отделима в системе S.

Пусть $\varphi_e(x)$ означает результат работы машины Тьюринга с номером e на входе x. Нетрудно убедиться, что функция

$$F(x) = \begin{cases} 1, & \text{если } \varphi_x(x) = 0, \\ 0, & \text{если } \varphi_x(x) \text{ определено и } \varphi_x(x) \neq 0, \\ \text{не определено} & \text{в остальных случаях} \end{cases} \tag{3}$$

вычислима, но не может иметь всюду определенного вычислимого продолжения. Действительно, допустим, что вычислимая функция $g = \varphi_n$ всюду определена. Рассмотрим $m = g(n) = \varphi_n(n)$. Если $m = 0$, то $F(n) = 1$, а если $m \neq 0$, то $F(n) = 0$. В любом случае $F(n) \neq m = g(n)$, т. е. g не продолжает F. Таким образом, неотделимые пары перечислимых множеств существуют.

Отсюда вытекает следующая абстрактная версия теоремы Гёделя–Россера.

Теорема 3. *Пусть формальная система $S = (L, P, R)$ непротиворечива и отделяет пары перечислимых множеств. Тогда S неотделима, и тем самым любое непротиворечивое расширение S неразрешимо и неполно.*

Так же как и понятие неотделимой пары перечислимых множеств, это утверждение восходит к работе Клини [14], который называл его “симметричной формой теоремы Гёделя”.

Замечание 1. Отметим, что теоремы 2 и 3 являются очень абстрактными и их применение для конкретных формальных систем, возникающих в логике, требует дополнительной работы (некоторые примеры будут рассмотрены ниже). Поэтому не следует буквально отождествлять эти утверждения с теоремами Гёделя и Россера.

Замечание 2. Существуют примеры неперечислимых теорий, для которых даже данные нами абстрактные формулировки не являются достаточно общими. Различные обобщения этих теорем в терминах так называемых *представляющих систем* очень подробно рассматривались Р. Смальяном [9]. Поскольку для нас центральной является вторая теорема Гёделя, мы не будем на этом подробно останавливаться.

3.2. Языки, теории, интерпретации.

Логические языки. Применения абстрактных результатов предыдущего пункта 3.1 прежде всего относятся к формальным системам, возникающим на базе логических языков, в частности, на базе языка логики предикатов первого порядка.

Мы будем рассматривать языки и теории первого порядка с равенством. Этот выбор, по существу, не является ограничительным, поскольку хорошо известно, что за счет расширения сигнатуры можно интерпретировать в таких языках и более богатые языки.

Прежде всего, можно интерпретировать многосортный язык, вводя в сигнатуру одноместные предикатные символы, выделяющие новые сорта переменных. Далее, вводя новый сорт переменных для функций f и дополнительную операцию $\mathrm{ap}(f,x)$, значением которой является результат $f(x)$ применения функции f к аргументу x, мы интерпретируем язык второго порядка с переменными по функциям. Аналогично мы можем интерпретировать переменные по предикатам, а также переменные высших порядков. Таким образом, излагаемые ниже результаты будут применимы к многосортным теориям и к теориям высших порядков. Этим покрывается значительная часть рассматриваемых на практике логических теорий, предназначенных для формализации математики.

Иногда рассматриваются некоторые фрагменты языка первого порядка, для которых также имеют место варианты теорем Гёделя. В частности, с этой точки зрения представляют интерес языки с ограниченными кванторами, бескванторные и эквациональные языки (языки равенств). Интерес к таким бедным языкам связан прежде всего с вопросом о том, насколько "простым" может быть гёделевское независимое утверждение. С другой стороны, имеется традиционный философский и исторический интерес к тематике программы Гильберта и формализации "финитной математики". В этом контексте бескванторные исчисления играют важную роль.

Теории. *Теориями* S мы будем называть теории первого порядка с равенством, т. е. теории, задаваемые некоторой сигнатурой Σ и множеством формул $\mathscr{P}$ данной сигнатуры, замкнутым относительно логического следования (в логике предикатов первого порядка с равенством). Формулы из $\mathscr{P}$ называем доказуемыми в S.

Теория $S = (\Sigma, \mathscr{P})$ называется *перечислимой*, если сигнатура Σ разрешима, а множество $\mathscr{P}$ перечислимо. Поскольку Σ разрешима, разрешимым будет также и множество всех формул $\mathscr{L}_\Sigma$ данного языка. Говоря о перечислимой теории S, мы будем предполагать фиксированными некоторые машины Тьюринга, задающие Σ и $\mathscr{P}$.

Как правило, машина Тьюринга, перечисляющая $\mathscr{P}$, определяется дедуктивным механизмом теории S. Например, если S задана конечным набором аксиом и правил вывода на базе гильбертовского исчисления предикатов, с таким заданием связывается определенный алгоритм перечисления всех возможных выводов (и выводимых формул) в S. Однако в принципе механизм перечисления теорем S может быть и совершенно другим: например, S может апеллировать к недедуктивным разрешающим процедурам для выяснения истинности тех или иных специальных классов формул (скажем, применять алгоритмы решения уравнений, если необходимость в этом возникает в процессе поиска вывода). Кстати, такого рода смешанные алгоритмы характерны для активно развивающихся сегодня компьютерных систем поиска вывода.

Известно также, что стандартный гильбертовский формат выводов во многих отношениях неудобен для практической работы с формальными выводами. В теории доказательств разработаны разнообразные альтернативные дедуктивные системы, которые удобны для тех или иных приложений: секвенциальные генценовские системы, системы так называемого натурального вывода и другие (см., например, [22]). Для первой теоремы Гёделя выбор конкретного дедуктивного механизма не играет большой роли.

Назовем *гёделевой нумерацией* формул языка $\mathscr{L}_\Sigma$ взаимно однозначное и вычислимое в обе стороны соответствие между множеством формул $\mathscr{L}_\Sigma$ и некоторым разрешимым подмножеством $L \subseteq \mathbb{N}$. Заметим, что, выбирая какую-либо гёделеву нумерацию формул языка $\mathscr{L}_\Sigma$, мы связываем с перечислимой теорией S некоторую формальную систему. (Считаем формулу φ опровержимой в S, если $\neg\varphi$ доказуемо в S. Поскольку операция $\varphi \mapsto \neg\varphi$ вычислима, множество опровержимых формул перечислимой теории также будет перечислимым.)

Формальная арифметика. Арифметика Пеано PA представляет собой теорию первого порядка с равенством в языке, содержащем константу 0 и символы для функций следования $S(x) = x + 1$, сложения $+$ и умножения $\cdot$. Стандартной моделью PA является множество натуральных чисел $\mathbb{N}$ (с нулем), рассматриваемое вместе со всеми этими операциями. Аксиомами PA, помимо логических аксиом и аксиом равенства, являются

P1. $\neg S(x) = 0$; $S(x) = S(y) \to x = y$;

P2. $x + 0 = x$; $x + S(y) = S(x + y)$;

P3. $x \cdot 0 = 0$; $x \cdot S(y) = (x \cdot y) + x$

вместе со *схемой аксиом индукции*

$$\varphi(0) \wedge \forall x \, \big(\varphi(x) \to \varphi(S(x))\big) \to \forall x \, \varphi(x)$$

для всех арифметических формул $\varphi(x)$ (возможно, содержащих параметры, т. е. свободные переменные помимо x).

Арифметика Робинсона Q получается заменой в формулировке PA схемы индукции на следующую аксиому (очевидно, выводимую в PA индукцией по x):

$$x = 0 \vee \exists y \, y = S(x).$$

Система Q традиционно играет большую роль в усилениях теорем Гёделя.

Интерпретации. Для языков и теорий первого порядка хорошо известно понятие относительной интерпретации (см., например, [23]). Интерпретации стали широко использоваться в математической логике, по-видимому, после работ Тарского и выхода монографии Тарского, Мостовского и Робинсона [21], где они активно применялись для исследования вопросов разрешимости логических теорий.

Теоремы Гёделя имеют место для языков и теорий в определенном смысле универсальных. Но для того, чтобы говорить об универсальности, нам необходимо иметь возможность сравнивать языки между собой, т. е. иметь дело с тем или иным понятием интерпретации. С этой точки зрения, достаточно общие формулировки теорем Гёделя естественным образом *предполагают* использование интерпретаций.[8]

[8] В большинстве изложений теорем Гёделя этот факт, тем не менее, остается за кадром или отмечается лишь на неформальном уровне.

Мы будем использовать достаточно широкое понятие интерпретации, которое можно описать как понятие многомерной интерпретации с определимыми параметрами и неабсолютным отношением равенства (точное определение см. в разделе 9). Такого рода интерпретации мы будем называть просто *интерпретациями*. Перевод формулы φ при интерпретации I обозначаем φ^I.

Когда мы фиксируем некоторую интерпретацию сигнатуры Σ_1 в Σ_2, мы фактически выделяем в языке Σ_2 новый сорт переменных, соответствующий объектам языка Σ_1, и предикаты и функции из Σ_1 над этими объектами (при этом выразительные способности языка Σ_2 не меняются). Поэтому часто мы будем рассуждать о языках и теориях как о многосортных.

Одним из центральных для нас является понятие *арифметической теории*. Арифметической теорией мы называем теорию, в которой фиксирована интерпретация исчисления предикатов в сигнатуре PA. Неформально мы можем считать, что в языке арифметической теории выделен особый сорт объектов – натуральные числа – вместе с обычными арифметическими операциями над этими объектами. Однако мы априори не предполагаем доказуемости каких-либо нелогических аксиом для этих символов.

Замечание 3. Забегая вперед, отметим, что, в принципе, теорию можно превратить в арифметическую теорию различными способами, что может влиять на ее метаматематические свойства. Рассмотрим две стандартные формализации теории множеств: Гёделя–Бернайса GB и Цермело–Френкеля ZF_g, где ZF_g рассматриваем в генценовском секвенциальном формате без правила сечения. Известно, что можно выбрать интерпретацию I арифметики Робинсона Q в теории множеств GB таким образом, чтобы для этой интерпретации было выполнено утверждение о непротиворечивости теории ZF_g, т. е. $\mathsf{GB} \vdash \mathrm{Con}(\mathsf{ZF}_g)^I$ (по этому поводу см., например, [24]). С другой стороны, для обычной интерпретации J натуральных чисел в теории множеств “по фон Нейману” это не так. В самом деле, GB консервативно расширяет ZF_g, и этот факт можно доказать в арифметике Пеано PA. Но аксиомы PA выполнены при фон-неймановской интерпретации натуральных чисел в GB, т. е. $\mathsf{GB} \vdash \mathsf{PA}^J$. Таким образом, в силу второй теоремы Гёделя $\mathsf{GB} \nvdash \mathrm{Con}(\mathsf{ZF}_g)^J$.

3.3. Σ_1-определимость. В языке арифметики неравенство $x \leqslant y$ обычно выражают как $\exists z\,(z + x = y)$. Добавим предикатный символ $\leqslant$ в сигнатуру арифметики и будем считать эквивалентность

$$x \leqslant y \quad \longleftrightarrow \quad \exists z\,(z + x = y)$$

еще одной аксиомой Q. В расширенном таким образом языке вводятся как сокращения *ограниченные кванторы*:

$$\forall x \leqslant t\,\varphi \quad \overset{\text{def}}{\Longleftrightarrow} \quad \forall x\,(x \leqslant t \to \varphi),$$
$$\exists x \leqslant t\,\varphi \quad \overset{\text{def}}{\Longleftrightarrow} \quad \exists x\,(x \leqslant t \wedge \varphi),$$

где терм t не содержит переменной x. Формула φ называется *ограниченной*, если любое вхождение квантора в φ ограничено, т. е. имеет вид $\forall x \leqslant t\,\psi$ или $\exists x \leqslant t\,\psi$. Множество всех ограниченных формул обозначается Δ_0.

Классы Σ_n- и Π_n-формул определяем индукцией по n следующим образом. Ограниченные формулы считаем как Σ_0-, так и Π_0-формулами. Σ_{n+1}-формулы суть формулы вида $\exists\vec{x}\,\varphi(\vec{x},\vec{y}\,)$, где φ есть Π_n-формула; Π_{n+1}-формулы суть формулы вида $\forall\vec{x}\,\varphi(\vec{x},\vec{y}\,)$, где φ есть Σ_n-формула.

Следующая фундаментальная теорема говорит о совпадении классов перечислимых и Σ_1-определимых множеств в стандартной модели $\mathbb{N}$.

Теорема 4. *Множество $A \subseteq \mathbb{N}^k$ перечислимо тогда и только тогда, когда для некоторой Σ_1-формулы $\varphi(\vec{x}\,)$*

$$\vec{n} \in A \iff \mathbb{N} \vDash \varphi[\vec{n}\,].$$

При этом формула φ строится эффективно (за полиномиальное время) по машине Тьюринга, задающей множество A, и наоборот.

Суть этой теоремы в том, что язык арифметических Σ_1-формул, интерпретируемых в $\mathbb{N}$, представляет собой универсальную модель вычислений. В смысле этой модели Σ_1-формула, определяющая график частичной функции, может рассматриваться как программа для вычисления этой функции, причем такая программа строится эффективно по машине Тьюринга, задающей данную функцию.

Следствие 1. *Множество всех истинных Σ_1-предложений перечислимо и неразрешимо. Множество всех истинных Π_1-предложений неперечислимо.*

Доказательство. Вторая часть следует из первой в силу теоремы Поста, поскольку истинные Π_1-предложения эффективно соответствуют ложным Σ_1-предложениям. Докажем первое утверждение.

Перечислимость множества истинных Σ_1-предложений $\mathrm{Th}_{\Sigma_1}(\mathbb{N})$ вытекает из того, что существует очевидный алгоритм проверки истинности данной Δ_0-формулы $\varphi(\vec{x}\,)$ на заданном наборе аргументов $\vec{n}$. Алгоритм, принимающий в точности истинные формулы вида $\exists\vec{x}\,\varphi(\vec{x}\,)$ для $\varphi \in \Delta_0$, состоит в переборе всевозможных аргументов $\vec{n}$ вплоть до первого набора, выполняющего φ.

Пусть теперь $K \subseteq \mathbb{N}$ – перечислимое, неразрешимое множество. Рассмотрим Σ_1-формулу φ_K, определяющую K в $\mathbb{N}$. Для каждого $n \in \mathbb{N}$ имеем:

$$n \in K \iff \mathbb{N} \vDash \varphi_K[n] \iff \mathbb{N} \vDash \varphi_K(\underline{n}\,).$$

Заметим, что Σ_1-формула $\varphi_K(\underline{n}\,)$ строится эффективно по n. Значит, вопрос о принадлежности $n \in K$ сводится к вопросу об истинности Σ_1-предложения $\varphi_K(\underline{n}\,)$. Следовательно, последний вопрос не может быть решен алгоритмически. Следствие 1 доказано.

Предшественником теоремы 4 был результат Гёделя об арифметичности примитивно рекурсивных отношений. Фактически, Гёдель указал способ построить по данной примитивно рекурсивной схеме некоторую арифметическую формулу, которую можно было бы назвать *обобщенной Σ_1-формулой.* С. Феферман [25] называет такие формулы RE-формулами. В формулах, получаемых конструкцией Гёделя, допускаются неограниченные кванторы существования в области действия ограниченного квантора общности. Однако можно заметить, что в стандартной модели арифметики выполняется *схема* Σ_1*-ограниченности*

$$\forall x \leqslant y\, \exists z\, \varphi(x,y,z) \quad \longleftrightarrow \quad \exists u\, \forall x \leqslant y\, \exists z \leqslant u\, \varphi(x,y,z),$$

что позволяет преобразовать обобщенную Σ_1-формулу в Σ_1-формулу. Тем самым, теорема 4 легко вытекает из первоначальной конструкции Гёделя.

Классы арифметических Σ_n- и Π_n-предикатов возникли в работах Клини [13] и Мостовского [19].

Понятие Δ_0-формулы и соответствующего класса предикатов на $\mathbb{N}$ было введено в книге Смальяна [16]. Для этого класса позже были получены интересные независимые характеризации, в частности, в терминах теории сложности вычислений. Класс Δ_0 в настоящее время играет важную роль в исследованиях по ограниченной арифметике (см., например, [26]). Использование Δ_0-определимых и Σ_1-определимых отношений вместо примитивно рекурсивных в доказательстве теоремы Гёделя позволяет избежать не нужного, по существу, формализма примитивно рекурсивных схем и работать напрямую в языке арифметики (см. [16]).

Хорошо известны усиления теоремы 4, связанные с возможным дальнейшим сужением класса Σ_1-формул. Так, известная теорема Матиясевича, опирающаяся на предшествующие результаты М. Дэвиса, Х. Патнема и Дж. Робинсон, давшая отрицательное решение 10-й проблемы Гильберта, утверждает, что всякое перечислимое отношение определимо даже формулой вида $\exists \vec{y}\, A(\vec{x}, \vec{y}\,)$, где A – равенство двух термов, т. е. многочленов с натуральными коэффициентами (см. [27], [28]). Такие формулы мы будем называть *диофантовыми*.

Доказательство теоремы 4 зависит от выбранной модели вычислений при определении перечислимого множества. Если используются машины Тьюринга, то задача сводится к построению Δ_0-определения предиката $T(e,x,y)$ "y кодирует протокол вычисления машины с номером e на входе x". Это сделать даже проще, чем арифметизировать по Гёделю предикат доказуемости в PA, поскольку правила работы машины Тьюринга более элементарны, чем синтаксис логики первого порядка.

Р. Смальян [16], используя некоторые технические находки Куайна, разработал один из самых простых методов арифметизации, который позволяет избежать использования β-функции Гёделя и китайской теоремы об остатках. Вместо машин Тьюринга он рассматривал другую удобную универсальную модель вычислений – так называемые *элементарные формальные системы*, родственные каноническим системам Поста.

3.4. Семантический вариант первой теоремы Гёделя. Как мы помним, формулируя свои теоремы, Гёдель избегал семантических понятий. Семантические варианты первой теоремы Гёделя, утверждающие существование при определенных условиях истинного недоказуемого утверждения, имеют более естественные формулировки и более простые доказательства. За эту простоту мы платим двумя вещами: во-первых, синтаксические варианты теоремы

Гёделя, как правило, сильнее семантических; во-вторых, мы используем неэлементарное понятие истинности в модели. В частности, мы даже не можем сформулировать семантический вариант теоремы Гёделя в языке арифметики.

Ниже мы увидим, что синтаксические варианты теоремы Гёделя получаются из семантических с помощью некоторой дополнительной технической идеи.

Напомним, что *арифметическими* мы называем теории, в которых фиксирована интерпретация языка арифметики Пеано. Можно считать, что язык арифметических теорий содержит выделенный сорт переменных по натуральным числам и все символы сигнатуры арифметики для данного сорта.

Определение 7. Пусть Γ – любое множество арифметических формул. Арифметическая теория T называется Γ-*полной*, если для любого предложения $A \in \Gamma$

$$\mathbb{N} \vDash A \implies T \vdash A.$$

T называется Γ-*корректной*, если верна обратная импликация: для любого предложения $A \in \Gamma$

$$T \vdash A \implies \mathbb{N} \vDash A.$$

Если Γ – множество всех арифметических предложений, теория T называется соответственно *семантически полной* и *корректной*.

Заметим, что Γ-полнота теории наследуется при переходе к ее расширениям, а Γ-корректность – при переходе к подтеориям. Поскольку все аксиомы PA истинны в стандартной модели, очевидно, PA корректна и тем самым Γ-корректна для любого Γ.

Лемма 1. *Пусть T – арифметическая теория.*

(i) *T является Σ_{n+1}-полной тогда и только тогда, когда она Π_n-полна.*

(ii) *T является Π_{n+1}-корректной тогда и только тогда, когда она Σ_n-корректна.*

Доказательство. (i) Если $\varphi \in \Pi_n$ и $\mathbb{N} \vDash \exists \vec{x}\, \varphi(\vec{x}\,)$, то $\mathbb{N} \vDash \varphi(\underline{\vec{n}}\,)$ для некоторого набора $\vec{n}$. Если T Π_n-полна, то $T \vdash \varphi(\underline{\vec{n}}\,)$, откуда $T \vdash \exists \vec{x} \varphi(\vec{x}\,)$ по правилам исчисления предикатов. Доказательство (ii) аналогично.

Теорема 4 позволяет получить первую теорему Гёделя о неполноте в следующей стандартной (семантической) формулировке.

Теорема 5. *Пусть T – перечислимая арифметическая теория.*

(i) *Если T непротиворечива, то T Π_1-неполна.*

(ii) *Если к тому же T Σ_1-корректна, то в T существует недоказуемое и неопровержимое Π_1-предложение.*

Доказательство. Допустим, что T Π_1-полна. Тогда, в силу предыдущей леммы, в T доказуемы все истинные и опровержимы все ложные Π_1-предложения (последние соответствуют истинным Σ_1-предложениям). Если T непротиворечива, то отсюда получаем для любого $\varphi \in \Pi_1$

$$\mathbb{N} \vDash \varphi \iff T \vdash \varphi.$$

В силу перечислимости T это означает, что множество истинных Π_1-предложений перечислимо, что противоречит следствию 1.

Допустим, что T Σ_1-корректна и φ – истинное недоказуемое в T Π_1-предложение. Тогда $\neg\varphi$ эквивалентно ложному Σ_1-предложению, откуда $T \nvdash \neg\varphi$. Теорема доказана.

Отметим, что в теореме 5 отсутствует бывшее первоначально у Гёделя предположение о том, что T содержит арифметические аксиомы. Теорема, в частности, применима и к чистой теории равенства в арифметическом языке. Однако, без дополнительных оговорок мы не можем утверждать, что теория T будет неразрешима.

Пример 1. Рассмотрим модель языка арифметики с носителем $\mathbb{N}$, в которой символы 0, S, $=$ имеют обычный смысл, а $x + y$ и $x \cdot y$ понимаются как $\max(x, y)$. Пусть T – элементарная теория этой модели. Разумеется, T некорректна (в смысле стандартной модели), но непротиворечива и синтаксически полна. Поскольку $\max$ выражается через $<$ и элементарная теория $(\mathbb{N}, <)$ разрешима, то T разрешима.

Этот пример показывает, что теорема 5 не является следствием абстрактной синтаксической версии первой теоремы Гёделя, данной выше.

3.5. Σ_1-полнота и синтаксический вариант теоремы Гёделя. Переход от семантического к синтаксическому варианту теоремы Гёделя опирается на свойство Σ_1-полноты, которое можно установить для широкого класса арифметических теорий.

Во-первых, конкретизируем понятие выразимости (определение 5) для случая арифметических теорий.

Определение 8. Множество $A \subseteq \mathbb{N}$ *представимо*[9] *в* T, если существует формула $\varphi_A(x)$ такая, что

$$n \in A \iff T \vdash \varphi_A(\underline{n}\,).$$

Заметим, что представимость множества A в T влечет его выразимость в T, поскольку функция подстановки $n \mapsto \varphi(\underline{n}\,)$ вычислима.

Лемма 2. *Если теория T является Σ_1-полной и Σ_1-корректной, то любое перечислимое множество представимо в T.*

Доказательство. Пусть A – перечислимое множество и φ_A – его Σ_1-определение в стандартной модели в смысле определения 4. Тогда, очевидно, φ_A представляет A в T. Лемма доказана.

Эта лемма позволяет применить данную нами выше абстрактную версию первой теоремы Гёделя, если установлена Σ_1-полнота теории T.

Отметим также, что для Σ_1-полных теорий семантическое условие Σ_1-корректности можно заменить на ослабленный вариант гёделевского условия ω-непротиворечивости.

---

[9]Это понятие называют также *нумеруемостью в* T, а формулу φ_A – нумерацией множества A в T [25].

Лемма 3. *Σ_1-полная арифметическая теория T является Σ_1-корректной тогда и только тогда, когда для любой формулы $\varphi(x) \in \Delta_0$ следующие условия не выполняются одновременно*:

(i) $T \vdash \exists x\, \varphi(x)$;

(ii) $T \vdash \neg\varphi(\underline{n})$ *для всех* $n \in \mathbb{N}$.

Условие, о котором идет речь в этой лемме, обычно называют 1-*непротиворечивостью* теории T.

Нам осталось найти достаточные условия Σ_1-полноты арифметической теории. Для этого достаточно указать одну, по возможности слабую, Σ_1-полную теорию. Традиционно в качестве такой теории фигурирует арифметика Робинсона Q или даже некоторая более слабая теория R (эта традиция восходит к монографии Тарского, Мостовского и Робинсона [21]). Теория Q сильнее R, но задается конечным числом аксиом.

Теорема 6. *Теория* Q *является* Σ_1*-полной.*

Идея доказательства Σ_1-полноты проста: истинность любого Δ_0-предложения φ может быть эффективно проверена. Такая проверка, по существу, представляет собой доказательство φ в Q, если φ истинно, или опровержение φ, если φ ложно. Фактически, для этого достаточно установить, что в теории Q доказуемы некоторые простейшие факты, касающиеся конкретных натуральных чисел.

Лемма 4. *Следующие формулы доказуемы в* Q:

R1. $\underline{m} + \underline{n} = \underline{m+n}$;

R2. $\underline{m} \cdot \underline{n} = \underline{m \cdot n}$;

R3. $\neg(\underline{m} = \underline{n})$, *если* $m \neq n$;

R4. $x \leqslant \underline{n} \;\leftrightarrow\; (x = 0 \vee x = \underline{1} \vee \cdots \vee x = \underline{n})$.

На основе этой леммы нетрудно установить полноту Q относительно Δ_0-предложений, откуда вытекает и Σ_1-полнота в силу леммы 1.

Теорию в языке арифметики, расширенном символом $\leqslant$, аксиоматизируемую всеми формулами вида R1–R4, называем *слабой арифметикой Робинсона* R_0. Таким образом, Σ_1-полнота имеет место и для любых расширений R_0. Отсюда получаем следующее следствие.

Следствие 2. *Пусть T – Σ_1-корректная теория, содержащая* Q *(или даже* R_0*). Тогда в T представимы все перечислимые отношения.*

Из теоремы 2 вытекает следующая теорема, которую можно считать стандартной синтаксической формулировкой первой теоремы Гёделя о неполноте.

Теорема 7. *Пусть T – арифметическая теория такая, что*

(i) *T содержит* Q *(или даже* R_0*)*,

(ii) *T перечислима*,

(iii) *T является Σ_1-корректной (или же 1-непротиворечивой).*

Тогда T неразрешима и неполна.

Отметим, что заключение этой теоремы сильнее второй части теоремы 5, поскольку дополнительно мы утверждаем и неразрешимость теории T.

Из теоремы 7 легко вывести вторую часть теоремы 5. Если T является Σ_1-корректной, то таковой является и расширение T всеми аксиомами R_0 (аксиомы R_0 бескванторны), а к этому расширению применима теорема 7.

Отметим также, что следствие 2 может быть установлено для любых непротиворечивых теорий T [8].

3.6. Теорема Гёделя–Россера. Абстрактная версия теоремы Гёделя–Россера опирается на свойство теории T отделять пары непересекающихся перечислимых множеств. Для доказательства этого свойства нам понадобится следующая лемма.

Лемма 5. *Для каждого $n \in \mathbb{N}$ в Q выводима формула*

$$\underline{n} \leqslant x \vee x \leqslant \underline{n}.$$

Расширение теории R_0 такими формулами для всех n обозначаем R, тем самым, R содержится в Q. Следующая лемма говорит о том, что теория R, а значит, и Q, отделяет пары непересекающихся перечислимых множеств.

Лемма 6. *Пусть $A, B \subseteq \mathbb{N}$ – непересекающиеся перечислимые множества. Тогда найдется Σ_1-формула $\varphi(x)$ такая, что для любого $n \in \mathbb{N}$*
(i) $n \in A \Rightarrow \mathsf{R} \vdash \varphi(\underline{n})$,
(ii) $n \in B \Rightarrow \mathsf{R} \vdash \neg\varphi(\underline{n})$.

Доказательство. По теореме 4 найдутся Δ_0-формулы A_0 и B_0 такие, что

$$\begin{aligned} n \in A &\iff \mathbb{N} \vDash \exists x\, A_0(\underline{n}, x), \\ n \in B &\iff \mathbb{N} \vDash \exists y\, B_0(\underline{n}, y). \end{aligned}$$

Для любой формулы C и терма t обозначим

$$\forall x < t\, C(x) \quad \stackrel{\text{def}}{\iff} \quad \forall x \leqslant t\, \big(x = t \vee C(x)\big).$$

Положим теперь

$$\varphi(z) \quad \stackrel{\text{def}}{\iff} \quad \exists x\, \big(A_0(z, x) \wedge \forall y < x \neg B_0(z, y)\big).$$

Неформально, $\varphi(z)$ утверждает, что работа алгоритма, принимающего множество A, на входе z заканчивается раньше работы алгоритма, принимающего B ("россеровское сравнение свидетелей").

Если $n \in A$, то для некоторого m истинна формула

$$A_0(\underline{n}, \underline{m}) \wedge \forall y < \underline{m} \neg B_0(\underline{n}, y).$$

В силу Σ_1-полноты арифметики R получаем, что эта формула доказуема в R, откуда $\mathsf{R} \vdash \varphi(\underline{n})$.

Если $n \in B$, то для некоторого m истинна формула

$$B_0(\underline{n}, \underline{m}) \wedge \forall y \leqslant \underline{m} \neg A_0(\underline{n}, y). \tag{4}$$

В силу Σ_1-полноты R получаем, что эта формула доказуема в R. Отсюда следует, что $\mathsf{R} \vdash \neg\varphi(\underline{n}\,)$. Поясним это следующим рассуждением, которое легко преобразовать в формальный вывод противоречия из гипотезы $\varphi(\underline{n}\,)$ в R.

Допустим $\varphi(\underline{n}\,)$. Тогда для некоторого x

$$A_0(\underline{n}, x) \wedge \forall y < x \neg B_0(\underline{n}, y).$$

Если $x \leqslant \underline{m}$, то имеем $\neg A_0(\underline{n}, x)$ из (4), что противоречит $A_0(\underline{n}, x)$. Если же $\underline{m} < x$, то имеем $\neg B_0(\underline{n}, \underline{m}\,)$, что противоречит $B_0(\underline{n}, \underline{m}\,)$ из (4). По лемме 4 в R выводимо $\forall x\,(x \leqslant \underline{m} \vee \underline{m} < x)$, откуда следует требуемое противоречие.

Лемма доказана.

Из этой леммы и теоремы 3 непосредственно получаем следующее утверждение.

Следствие 3. (i) *Система* R *неотделима.*

(ii) *Любая непротиворечивая арифметическая теория, содержащая* R, *неразрешима.*

Отсюда также вытекает *теорема Гёделя–Россера.*

Теорема 8. *Пусть T – арифметическая теория такая, что*

(i) *T содержит* R,

(ii) *T перечислима,*

(iii) *T непротиворечива.*

Тогда T неотделима и, тем самым, неразрешима и неполна.

3.7. Эффективность теорем Гёделя и Россера. На первый взгляд, приведенные доказательства теорем Гёделя и Россера не позволяют получить явные примеры независимых арифметических утверждений, поскольку опираются на рассуждения от противного. Однако эти доказательства легко модифицировать таким образом, чтобы это стало возможно.

Напомним, что стандартное перечислимое неразрешимое множество

$$K = \{x \in \mathbb{N} : \varphi_x(x) \text{ определено}\}$$

является *креативным*, т. е. для любого перечислимого множества

$$W_n = \{x \in \mathbb{N} : \varphi_n(x) \text{ определено}\}$$

можно эффективно по n указать число x такое, что $x \notin K \cup W_n$, как только $K \cap W_n = \varnothing$. Фактически, в качестве x можно взять само n, поскольку $n \in K \iff n \in W_n$.

Пусть $\psi_K(x)$ – Σ_1-формула, нумерующая K в T (или выражающая K в стандартной модели арифметики). Доказательства теорем 5, 7 опирались на то, что в предположении полноты T формула $\neg\psi_K(x)$ нумерует дополнение множества K. Однако без этого (ложного) предположения мы можем лишь утверждать, что $\neg\psi_K(x)$ нумерует некоторое перечислимое подмножество K' дополнения K. Некоторый индекс этого множества, т. е. число n, для которого $K' = W_n$, может быть найден явно, поскольку нам известны как формула

$\neg\psi_K(x)$, так и программа, перечисляющая теоремы T. Используя креативность множества K, получаем число m, для которого $m \notin K \cup K'$, откуда следует, что ни $\psi_K(\underline{m})$, ни $\neg\psi_K(\underline{m})$ не доказуема в T. Таким образом, использование креативности K позволяет получить эффективный вариант первой теоремы Гёделя.

Для анализа теоремы Гёделя–Россера мы используем эффективный вариант понятия неотделимой пары перечислимых множеств. Как мы помним, пары непересекающихся перечислимых множеств отождествляются с вычислимыми частичными функциями $g\colon \mathbb{N} \to \{0,1\}$. Стандартный пример неотделимой пары перечислимых множеств дает вычислимая функция F из (3), не продолжаемая до всюду определенной вычислимой функции. Эта функция фактически обладает более сильным свойством *эффективной непродолжаемости*: по всякой вычислимой функции φ_n, продолжающей F, можно эффективно по n указать такое m, что $\varphi_n(m)$ не определено.[10] (Соответствующую пару перечислимых множеств называют *эффективно неотделимой*.)

Доказательство теоремы Гёделя–Россера базируется на лемме 5, которая эффективна в том смысле, что позволяет по заданной паре перечислимых множеств (A, B) явно указать формулу ψ, отделяющую пару (A, B) в теории R. В нашем случае мы рассматриваем пару (A, B), заданную машиной Тьюринга M, вычисляющей эффективно непродолжаемую частичную функцию F. Пользуясь эффективностью построения Σ_1-определений множеств A и B по машине M, получаем соответствующую формулу ψ_M.

С формулой ψ_M связывается теперь другая вычислимая частичная функция

$$f(n) = \begin{cases} 1, & \text{если } T \vdash \psi_M(\underline{n}), \\ 0, & \text{если } T \vdash \neg\psi_M(\underline{n}), \\ \text{не определено} & \text{в остальных случаях.} \end{cases}$$

По определению ψ_M мы знаем, что функция f продолжает F. Заметим, что по формуле ψ_M и программе, перечисляющей теоремы T, мы можем эффективно найти индекс n функции f. Эффективная непродолжаемость F дает число m, для которого $f(\underline{m})$ не определено, т. е. ни $\psi_M(\underline{m})$, ни $\neg\psi_M(\underline{m})$ не доказуема в T.

4. О границах применимости первой теоремы Гёделя

Как семантическая, так и синтаксическая формы теоремы Гёделя содержат несколько условий. Анализировать их удобно, зафиксировав некоторые из них и варьируя остальные. Наиболее важным условием является перечислимость теории T, поэтому мы рассмотрим роль других условий в предположении перечислимости T.

Заметим, что при этом предположении факт неполноты теории T выводится из более сильного факта ее неразрешимости. Тем самым, исследование вопроса о неполноте теории T превращается в исследование условий, при которых гарантируется ее неразрешимость.

[10] Для данной F фактически $m = n$.

Вопросы разрешимости и неразрешимости логических теорий являются одними из центральных в математической логике. В частности, им уделялось большое внимание в работах А. И. Мальцева и его школы. В настоящей статье нет возможности дать сколько-нибудь полный обзор этой темы. Укажем лишь монографию [29] и замечательный обзор [30], до сих пор не утративший своей актуальности. Здесь мы отметим лишь некоторые отдельные направления исследований, наиболее близкие к рассматриваемой нами задаче.

Семантическая формулировка теоремы Гёделя для перечислимых теорий сводится к утверждению о неперечислимости множества истинных арифметических Π_1-предложений. Возможные вариации этой теоремы идут по двум направлениям.

Во-первых, оставаясь в рамках языка арифметики, можно исследовать более узкие классы формул, чем Π_1, для которых сохраняется неперечислимость. Смысл таких исследований состоит в поиске наиболее простых по форме неразрешимых задач и, соответственно, независимых предложений. Самые интересные результаты в этом направлении связаны с исследованиями 10-й проблемы Гильберта и с так называемыми диофантовыми формами теоремы Гёделя (более подробно об этом см. ниже).

Во-вторых, очень активно изучались вопросы о разрешимости теорий в других языках; причем особенно подробно исследовались разрешимые фрагменты языка арифметики. Начиная с классических результатов М. Пресбургера и Т. Сколема, показавших разрешимость элементарных теорий $(\mathbb{N};=,+)$ и $(\mathbb{N};=,\cdot)$ соответственно, важный вклад в это направление внесли работы Дж. Робинсон, Й. Бюхи, М. Рабина, А. Вудса, Ю. В. Матиясевича, А. Л. Семёнова, Ан. А. Мучника, Д. Ришара, П. Сегиельски и других. Современный обзор результатов по этой теме дан в работе [31].

Синтаксическая формулировка теоремы Гёделя (в более сильной форме Россера) является непосредственным следствием другого утверждения – неразрешимости любого непротиворечивого расширения системы R (следствие 3). Теории, все непротиворечивые расширения которых неразрешимы, называются *существенно неразрешимыми.*

Тарский [21] заметил, что свойство непополняемости перечислимой теории T не только следует из условия ее существенной неразрешимости, но и равносильно ему. Действительно, эффективный вариант леммы Линденбаума показывает, что всякая разрешимая теория имеет некоторое разрешимое пополнение. Таким образом, исследование условий непополняемости перечислимых теорий превращается в исследование условий их существенной неразрешимости. Заметим, что выше было доказано более сильное свойство неотделимости теории R (следствие 3).

Тарским было найдено следующее достаточное условие неразрешимости теорий, использующее конечно аксиоматизируемые существенно неразрешимые теории, такие как Q.

Предложение 1. *Пусть S – конечно аксиоматизируемая, существенно неразрешимая теория. Тогда любая совместная с S теория T той же сигнатуры неразрешима.*

Доказательство. Допустим, что T разрешима. Рассмотрим непротиворечивую теорию $T' = T + S$. Поскольку S задана конечным числом аксиом, мы имеем

$$T' \vdash \varphi \iff T \vdash A_S \to \varphi,$$

где A_S есть конъюнкция универсальных замыканий аксиом S. Таким образом, проблема выводимости в T' сводится к таковой в T, и, следовательно, T' также разрешима, что противоречит существенной неразрешимости S. Предложение доказано.

Заметим, что отсюда, в частности, вытекает теорема Чёрча о неразрешимости исчисления предикатов (в сигнатуре арифметики). Как показал Тарский, это предложение имеет очень полезное усиление в терминах интерпретаций.

Определение 9. Говорим, что теория S *слабо интерпретируема* в T, если S интерпретируема в некоторой совместной с T теории U в языке T.

Предложение 2. *Пусть S – конечно аксиоматизируемая, существенно неразрешимая теория. Если S слабо интерпретируема в теории T, то T (и любая ее подтеория в том же языке) неразрешима.*

Доказательство. Пусть I – интерпретация S в теории U, совместной с T. Считаем, что параметры I определяются некоторой формулой $\mathrm{Par}_I(\vec{p})$. Можно также считать, что S задается единственным предложением A_S, тогда в U доказуемо

$$\forall \vec{p}\,\bigl(\mathrm{Par}_I(\vec{p}) \to A_S^I(\vec{p})\bigr).$$

Пусть теория U_0 задается этим предложением, тогда U_0 совместна с T. Но U_0 также существенно неразрешима: если $U_0 \subseteq V$ и теория V непротиворечива, то множество

$$\bigl\{\varphi : V \vdash \forall \vec{p}\,\bigl(\mathrm{Par}_I(\vec{p}) \to \varphi^I(\vec{p})\bigr)\bigr\}$$

дедуктивно замкнуто, непротиворечиво и содержит S. Осталось применить предложение 1. Предложение 2 доказано.

Таким образом, для доказательства неразрешимости теорий методом интерпретаций становится важным получить примеры слабых конечно аксиоматизируемых существенно неразрешимых теорий. Тарский, Мостовский и Робинсон использовали для этих целей теорию Q. Формально ослабить Q без потери существенной неразрешимости можно за счет замены функций $+$ и $\cdot$ на трехместные предикаты $A(x,y,z)$ и $M(x,y,z)$. Соответствующая система Q^- была не так давно сформулирована А. Гжегорчиком (см. [32]). Ее аксиомами являются три аксиомы Q, касающиеся функции следования, а также варианты остальных аксиом в реляционном языке, не предполагающие, что функции сложения и умножения всюду определены:

1. $A(x,y,u) \wedge A(x,y,v) \;\to\; u = v$; $M(x,y,u) \wedge M(x,y,v) \;\to\; u = v$;
2. $A(x,0,x)$; $\exists u\,\bigl(A(x,y,u) \wedge z = S(u)\bigr) \;\to\; A\bigl(x,S(y),z\bigr)$;
3. $M(x,0,0)$; $\exists u\,\bigl(M(x,y,u) \wedge A(u,x,z)\bigr) \;\to\; M\bigl(x,S(y),z\bigr)$.

В. Швейдар [32] доказал, что теория Q интерпретируема в Q^-, поэтому Q^- существенно неразрешима. Однако дедуктивно Q^- слабее, чем Q, и даже не является Σ_1-полной (с учетом аксиомы, определяющей $\leqslant$ через сложение); в частности, в Q доказуемо $\forall x \leqslant 0\, x = 0$, а в Q^- нет.

Слабые конечно аксиоматизированные существенно неразрешимые теории известны для языка конкатенации слов в двоичном алфавите, для слабой теории множеств и др. (см. [33]–[36]). Например, ослабив теорию Тарского и Шмелевой, и тем самым усилив их результат, Р. Воот показал, что таковым является вариант теории множеств в сигнатуре $\in$ лишь с двумя аксиомами:

$$\forall x\, \exists y\, \neg(y \in x), \qquad \forall x, y\, \exists u\, \forall z\, \bigl(z \in u \leftrightarrow (z \in x \vee z \in y)\bigr).$$

До настоящего времени эта теория, по-видимому, является самым простым по формулировке примером существенно неразрешимой конечно аксиоматизируемой теории.

Х. Патнэм и А. Эренфойхт [37], [38] построили примеры перечислимых существенно неразрешимых теорий S, для которых предложение 2 не имеет места. На этом фоне представляет интерес результат А. Кобхэма (см. [33]) о том, что предложение 2 верно и для таких теорий, как R и даже R_0, не являющихся конечно аксиоматизируемыми. Обсудим этот результат.

Во-первых, отметим, что наша формулировка системы R несколько отличается от традиционной из [21]. В традиционной формулировке знак $\leqslant$ вводится как сокращение и аксиома R4 ослаблена до импликации слева направо. Однако легко видеть, что R4 является выводимой и в традиционной формулировке R.

Во-вторых, имеет место следующее предложение (см. [39]).

Предложение 3. (i) R *интерпретируема в* R_0.

(ii) R_0 *существенно неразрешима и неотделима.*

Доказательство. Определим интерпретацию I теории R в R_0, видоизменив отношение порядка в R_0 и оставив все остальные символы сигнатуры без изменений. Положим

$$x \leqslant_I y \ \stackrel{\text{def}}{\iff}\ \bigl[\bigl(0 \leqslant y \wedge \forall u\, (u \leqslant y \wedge u \neq y \to S(u) \leqslant y)\bigr) \to x \leqslant y\bigr].$$

Легко видеть, что для отношения $\leqslant_I$ в R_0 будут доказуема схема R4, а также дополнительная схема $x \leqslant_I \underline{n} \vee \underline{n} \leqslant_I x$. Тем самым установлено (i).

Интерпретация I не содержит параметров. Следовательно, для любого предложения φ

$$\mathsf{R} \vdash \varphi \implies \mathsf{R}_0 \vdash \varphi^I.$$

Поскольку I сохраняет нумералы, отсюда непосредственно вытекает, что R_0, так же как и R, отделяет пары перечислимых множеств и тем самым является неотделимой теорией. Предложение доказано.

Следствие 4. *Теорема Гёделя–Россера имеет место и для непротиворечивых расширений* R_0.

Теорию R_0 можно и далее ослаблять. Используя идею Дж. Робинсон, показавшей, что в модели $(\mathbb{N}; =, S, \cdot)$ выразимо сложение, Дж. Джонс и Дж. Шепердсон [39] установили интерпретируемость R_0 в теории R_1, получающейся из R_0 отбрасыванием операции сложения и схемы R1. Отсюда, как и ранее, следует как существенная неразрешимость, так и неотделимость R_1.

С другой стороны, сам Кобхэм доказал свою теорему для более слабого реляционного варианта теории R_0, который мы обозначим R_0^-.

Язык R_0^- содержит предикатные символы C_0, Sc, A, M, $\leqslant$ валентностей 1, 2, 3, 3, 2 соответственно и не содержит равенства. Предикат $C_0(x)$ выделяет константу 0, а Sc, A и M определяют графики функций следования, сложения и умножения. Вместо нумералов используем формулы $C_n(x)$, индуктивно определяемые как $C_{n+1}(x) \stackrel{\text{def}}{\Longleftrightarrow} \exists y\, \big(C_n(y) \wedge Sc(y, x)\big)$.

Теория R_0^- содержит следующие аксиомы [11]:

(i) $\exists x\, C_0(x)$; $\neg\big(C_m(x) \wedge C_n(x)\big)$ для $m \neq n$;

(ii) $C_m(x) \wedge C_n(x) \ \to\ \big(A(x, y, z) \leftrightarrow C_{m+n}(z)\big)$;

(iii) $C_m(x) \wedge C_n(x) \ \to\ \big(M(x, y, z) \leftrightarrow C_{mn}(z)\big)$;

(iv) $C_n(y) \ \to\ \big(x \leqslant y \leftrightarrow C_0(x) \vee C_1(x) \vee \cdots \vee C_n(x)\big)$.

Заметим, что R_0^- не содержит аксиом равенства, поэтому ощутимо слабее R_0 и удобнее в применениях, основу которых составляет следующая теорема Кобхэма.

Теорема 9. *Если R_0^- слабо интерпретируема в T, то T (и любая ее подтеория) неразрешима.*

Существенная неразрешимость R_0^- непосредственно следует из данной теоремы. Однако ее достаточно просто установить и непосредственно.

Поскольку язык R_0^- не содержит нумералов, понятие разрешимости k-местного отношения A в теории T сигнатуры R_0^- модифицируется следующим образом: существует формула $\varphi_A(x_1, \ldots, x_k)$ такая, что для всех наборов чисел $n_1, \ldots, n_k$

$$A(n_1, \ldots, n_k) \implies T \vdash C_{n_1}(x_1) \wedge \cdots \wedge C_{n_k}(x_k) \ \to\ \varphi_A, \qquad (5)$$

$$\neg A(n_1, \ldots, n_k) \implies T \vdash C_{n_1}(x_1) \wedge \cdots \wedge C_{n_k}(x_k) \ \to\ \neg\varphi_A. \qquad (6)$$

С учетом этой модификации нетрудно доказать разрешимость в R_0^- любых Δ_0-предикатов, так же как и отделимость пар перечислимых множеств.

Следствие 5. *Теория R_0^- неотделима и существенно неразрешима.*

Из теоремы Кобхэма непосредственно вытекает одна из наиболее сильных форм теоремы Гёделя–Россера.

Теорема 10. *Если R_0^- слабо интерпретируема в перечислимой теории T, то T неразрешима и неполна.*

[11] В определении R_0^- в [33] и [30] допущена опечатка: в аксиомах (ii) и (iii) вместо $\leftrightarrow$ стоит $\to$. Такая теория не может быть существенно неразрешимой, поскольку она интерпретируется в разрешимой теории $\mathrm{Th}(\mathbb{N}; \leqslant)$ при переводе A и M тождественно ложными формулами.

Р. Воот [33] предложил достаточно простой вывод теоремы Кобхэма 9 из теоремы Трахтенброта о неотделимости множества тождественно истинных формул логики предикатов и множества формул, опровержимых на конечных моделях. (В качестве сигнатуры при этом достаточно взять один бинарный предикатный символ.) Таким образом, теорема Трахтенброта дает еще один способ доказать теорему Гёделя–Россера.

А. Виссер [40] дал интересную характеризацию теорий, взаимно интерпретируемых с R (или R_0), в терминах выполнимости на конечных моделях.

Назовем теорию T *конечно выполнимой*, если T имеет конечную модель. Теория T *локально конечно выполнима*, если конечно выполнима любая конечная подтеория T. Очевидно, что R является локально конечно выполнимой, но не конечно выполнимой теорией. Виссер показал, что R обладает следующим свойством максимальности.

Теорема 11. *Всякая перечислимая, локально конечно выполнимая теория T интерпретируема в* R (*причем интерпретация одномерная и без параметров*).

Эта теорема показывает, что если в выборе сигнатуры, аксиом и некоторых других деталей в формулировке R и есть произвол, то по модулю интерпретируемости R занимает привилегированную позицию и выделяется некоторым естественным общим свойством.

Отметим также, что помимо арифметических теорий типа R рассматривались и существенно неразрешимые локально конечно выполнимые теории в других языках. Одним из самых красивых примеров является теория S, введенная Воотом [41]. S задается в сигнатуре теории множеств единственной схемой аксиом

$$\forall x_1, \ldots, x_n\, \exists y\, \forall z \left(z \in y \leftrightarrow \bigvee_{i=1}^{n} z = x_i \right) \quad \text{для всех } n \geqslant 1.$$

Воот показал, что S интерпретирует R_0^- и тем самым является неотделимой и существенно неразрешимой.

5. О доказательствах первой теоремы Гёделя

Многие встречающиеся в литературе доказательства первой теоремы Гёделя следуют изложенной нами общей схеме, однако могут значительно отличаться друг от друга в технических деталях. Восстановление всех деталей оставляет на вкус авторов как минимум выбор следующих параметров (которые мы будем обсуждать для более простой семантической версии теоремы Гёделя):

1. Выбор конкретной базисной формальной системы T;

2. Выбор универсальной модели вычислений на некотором семействе U объектов T; под такой моделью я понимаю любое математически строгое определение понятия перечислимого множества элементов U (или же вычислимой функции из U в U);

3. Выбор гёделевой нумерации, т. е. кодирование синтаксиса теории T объектами из U;

4. Доказательство перечислимости системы T (в смысле выбранной модели вычислений и гёделевой нумерации);

5. Предъявление примера выразимого неперечислимого множества (вместе с доказательством его выразимости и неперечислимости).

Отметим сразу, что многие авторы упрощенных доказательств теоремы Гёделя игнорируют некоторые из этих пунктов за их интуитивной очевидностью, пользуясь при этом, как правило, неформальным представлением об алгоритмах и какой-либо из форм тезиса Чёрча–Тьюринга. Например, интуитивно очевидной является перечислимость арифметики PA. В то же время, честное обоснование этого утверждения требует программирования в рамках выбранной модели вычислений, т. е., вообще говоря, значительной технической работы. Выбрав аппарат примитивно рекурсивных функций, Гёдель вполне эффективно справился с этой задачей. Разумеется сравнение может быть справедливым лишь для более или менее *полных* доказательств, хотя с педагогической точки зрения игнорирование интуитивно очевидных утверждений вполне оправдано и, более того, весьма желательно. На наш взгляд, существенно более простых, чем у Гёделя, полных доказательств его теоремы все-таки нет.

Разберем некоторые из названных пунктов по очереди.

Выбор теории. В качестве системы T, как правило, однако не всегда, выбирают некоторую арифметическую теорию. В педагогических или идеологических целях здесь также возможны различные варианты в зависимости от выбора сигнатуры. Например, удобно развивать кодирование в языке, имеющем символ для функции возведения в степень.

Часто в качестве T рассматривают примитивно рекурсивную арифметику PRA, в языке которой имеются термы для всех примитивно рекурсивных функций. В такой системе можно напрямую работать с примитивно рекурсивными функциями и избежать неудобств их погружения в язык PA [3]. Язык PRA бесконечен, хотя и примитивно рекурсивен. Кроме того, PRA является слишком сильной теорией, далеко за пределами теорий ограниченной арифметики, что затрудняет обобщение теоремы Гёделя на этот важный класс формальных систем.

В качестве альтернативы арифметическим теориям рассматриваются также теории помеченных бинарных деревьев [42], теории слов в бинарном алфавите [43], [44], теории наследственно конечных множеств [45] и другие. Описание синтаксиса формальных теорий или же моделей вычислений в таких системах более естественно, так как позволяет в некотором смысле избежать гёделевой нумерации: формулы уже сами по себе отождествляются с объектами теории (словами, помеченными деревьями или же конечными множествами).

Возможны вариации и в выборе дедуктивной системы T: можно рассматривать натуральный вывод, генценовское секвенциальное исчисление и многие другие форматы выводов. Как уже отмечалось, для первой теоремы Гёделя эти вариации не столь принципиальны и, как правило, выбирают наиболее привычный и просто формулируемый гильбертовский формат.

Выбор модели вычислений. В качестве модели вычислений для доказательства теоремы Гёделя разными авторами рассматривались:

1) перечислимые множества как проекции примитивно рекурсивных отношений (К. Гёдель);
2) функции, вычислимые по Эрбрану–Гёделю, частично рекурсивные функции (С. Клини);
3) элементарные формальные системы (Р. Смальян);
4) машины Тьюринга;
5) Σ-определимые отношения (Ю. Л. Ершов) и другие.

Выбор каждой из этих моделей имеет свои преимущества и недостатки.

Машины Тьюринга, с моей точки зрения, хороши тем, что являются наиболее стандартной и естественной моделью вычислений. Из всех известных теоретических моделей именно для них тезис Чёрча–Тьюринга наиболее убедителен. По этой причине они идеально подходят для "сокращенных" доказательств теоремы Гёделя.

Поскольку на машинах Тьюринга несколько труднее программировать, чем на более высокоуровневом языке частично рекурсивных функций, явное построение универсального перечислимого множества, так же как и доказательство перечислимости арифметики, будет не столь простым, как на языке частично рекурсивных функций.

Интересен выбор Σ-определимости как самостоятельной модели вычислений (Ершов [45]). При этом теорема 4 превращается, по сути, в определение перечислимого множества и не требует доказательства. В качестве расплаты за такой выбор возникает необходимость построения универсального перечислимого множества в рамках данной модели. Нетрудно видеть, что таким множеством является та формула, которую в логике принято называть Σ-определением истинности для Σ-формул (см. [26], [45]). Задача построения этой формулы по своей технической трудности аналогична задаче арифметического определения предиката доказуемости. Тем не менее, концептуально выбор Σ-определимости как модели вычислений хорош тем, что все доказательство проходит в рамках одного языка – будь то теория множеств или формальная арифметика – и не требуется привлечение каких-то внешних механизмов вычислимости.

Из известных "честных" доказательств теоремы Гёделя одним из образцовых в смысле продуманности дизайна и деталей является доказательство Р. Смальяна [16], [9]. В частности, его выбор элементарных формальных систем как модели вычислимости прекрасно адаптирован к потребностям формализации логических языков.

Выбор неперечислимого множества. Наиболее радикально влияет на внешнюю сторону доказательства – не меняя, разумеется, сути дела – выбор конкретного неперечислимого множества, используемого для построения независимого утверждения. Иногда такое множество связывается с формальным аналогом какого-нибудь семантического парадокса. Доказательство самого Гёделя (см. ниже) основывается на построении формулы, утверждающей собственную недоказуемость, что аналогично парадоксу лжеца. (Заметим, что если теория T семантически полна, то утверждение о недоказуемости формулы φ в T равносильно ложности φ.) Однако Гёдель отметил, что для доказательства его теорем может быть приспособлен почти любой из известных семантических парадоксов. Многие позднейшие авторы с успехом подтвердили эту мысль классика.

Клини [12] был, по-видимому, первым, кто изложил абстрактную вычислительную канву теоремы Гёделя, в духе обсуждаемой здесь, на основе понятия общерекурсивной функции по Эрбрану–Гёделю. Для своего первого теоретико-рекурсивного доказательства теоремы Гёделя он воспользовался неперечислимостью множества индексов общерекурсивных (т. е. всюду определенных вычислимых) функций. Как он сам отметил позже [13], это доказательство наиболее близко парадоксу Ришара (и диагональной конструкции Кантора).

Другие известные доказательства связаны с парадоксом Берри: существует ли "*наименьшее натуральное число, которое нельзя определить менее чем пятнадцатью словами русского языка*"? Два доказательства, использующие похожую идею, были предложены Г. Чейтином [46] и Дж. Булосом [47]. Булос в качестве уточнения понятия "определить" использует обычное понятие определимости в арифметике. На доказательстве Чейтина остановимся чуть подробнее, поскольку оно приобрело большую известность, в том числе в научно-популярной литературе.

Доказательство Чейтина базируется на понятии *колмогоровской сложности* числа x, обозначаемой $K(x)$, т. е. в качестве "определений" x рассматриваются всевозможные программы p такие, что на пустом входе p выдает x. В знакомых нам терминах мы можем определить один из вариантов функции K для стандартной (оптимальной) нумерации машин Тьюринга как

$$K(x) = \min\{n \in \mathbb{N} : \varphi_n(0) = x\}.$$

В силу того, что отношение $K(x) > y$ неперечислимо и выразимо в арифметике, мы можем заключить, что существуют недоказуемые истинные утверждения вида $K(\underline{n}\,) > \underline{c}$ для некоторых констант n, c. Чейтин утверждает нечто большее, а именно предъявляет целую серию недоказуемых утверждений такого вида.

Теорема 12. *Пусть T – перечислимая и корректная арифметическая теория. Тогда*

$$\exists c\, \forall m\, T \nvdash K(\underline{m}\,) > \underline{c}.$$

Вычислительная основа этой теоремы состоит в следующем хорошо известном свойстве колмогоровской сложности: не существует эффективного способа по c найти число m, для которого $K(m) > c$. (В то же время, очевидно, что для всякого c такое m существует, поскольку число программ с номерами, не превосходящими c, конечно.)

Лемма 7. *Не существует всюду определенной вычислимой функции f такой, что $\forall c\, K(f(c)) > c$.*

Теорема Чейтина следует из этой леммы: если

$$\forall c\, \exists m\, T \vdash K(\underline{m}\,) > \underline{c},$$

то в силу перечислимости T получаем вычислимую всюду определенную f, для которой $\forall c\, T \vdash K(\underline{f(c)}) > \underline{c}$, а в силу корректности T отсюда следует $\forall c\, K(f(c)) > c$ в противоречии с леммой.

Для доказательства леммы воспользуемся теоремой Клини о рекурсии. Рассуждая от противного, допустим, что существует f, удовлетворяющая условиям леммы. Рассмотрим вычислимую функцию g, которая по входу c сначала вычисляет $y = f(c)$, а затем выдает некоторый номер n такой, что $\varphi_n(0) = y$. Пользуясь теоремой о рекурсии, выберем индекс e такой, что функция φ_e совпадает с $\varphi_{g(e)}$. По построению g имеем $K(f(e)) \leqslant g(e)$, но, поскольку программа e делает то же, что и $g(e)$, мы также имеем $K(f(e)) \leqslant e$, что противоречит условию леммы.

Приведенное нами короткое доказательство теоремы Чейтина с использованием теоремы о рекурсии есть переработанный вариант из статьи [48], см. также [49]. Сам Чейтин дал другое, более длинное, но прямое, доказательство (см. [46], [50], а также [51]). Отметим, что константа c зависит от выбора рассматриваемой нумерации вычислимых функций и при некотором выборе такой нумерации можно добиться равенства c нулю для всех теорий [48].

Как мы видим, теорема Чейтина является непосредственным проявлением невычислимости некоторых простых свойств, связанных с функцией колмогоровской сложности $K(x)$. Такого рода факты о $K(x)$ были, разумеется, хорошо известны, так же как и общая схема, позволяющая их конвертировать в результаты о неполноте. С этой точки зрения результат Чейтина не более интересен, чем другие доказательства теоремы о неполноте, следующие изложенной выше общей схеме.

6. Доказательство Гёделя

С точки зрения теории вычислимости наиболее интересен, пожалуй, анализ первоначального доказательства самого Гёделя и его позднейших версий, скажем, для теоремы 7 (см., например, [52], [15]). Это доказательство основывается на техническом понятии *представимости функции* в теории T и построении арифметической формулы, утверждающей собственную недоказуемость. Общую схему можно описать так:

1) доказательство примитивной рекурсивности предиката доказательств $\mathrm{Prf}_T(x,y)$ для теории T;
2) доказательство представимости всех примитивно рекурсивных функций в T, откуда следует разрешимость в T предиката $\mathrm{Prf}_T(x,y)$;
3) построение формулы ψ такой, что

$$T \vdash \psi \leftrightarrow \neg\mathrm{Pr}_T(\underline{\ulcorner\psi\urcorner}),$$

где $\mathrm{Pr}_T(x)$ означает формулу доказуемости $\exists y\, \mathrm{Prf}_T(x,y)$; при этом мы используем представимость в T функции подстановки.

Доказательство завершается следующим рассуждением, показывающим, что при условии ω-непротиворечивости T формула ψ недоказуема и неопровержима в T.

Если $T \vdash \psi$, то $\mathbb{N} \vDash \mathrm{Prf}_T\big(\ulcorner\psi\urcorner, n\big)$ для некоторого n, откуда $T \vdash \mathrm{Prf}_T\big(\underline{\ulcorner\psi\urcorner}, \underline{n}\,\big)$ в силу разрешимости Prf_T в T, и тем самым $T \vdash \mathrm{Pr}_T\big(\underline{\ulcorner\psi\urcorner}\big)$. По определению ψ это влечет $T \vdash \neg\psi$, т. е. противоречивость T.

Если $T \vdash \neg\psi$, то $T \vdash \exists y\, \mathrm{Prf}_T\big(\underline{\ulcorner\psi\urcorner}, y\big)$ по определению ψ. С другой стороны, в силу непротиворечивости T имеем $T \nvdash \psi$. Значит, для каждого конкретного n выполнено $\mathbb{N} \vDash \neg\mathrm{Prf}_T\big(\ulcorner\psi\urcorner, n\big)$. Поскольку предикат Prf_T разрешим в T, получаем, что для каждого n $T \vdash \neg\mathrm{Prf}_T\big(\underline{\ulcorner\psi\urcorner}, \underline{n}\,\big)$, т. е. T является ω-противоречивой.

Отметим, что шаг 1) доказательства представляет собой один из исторически первых опытов программирования, в данном случае на языке примитивно рекурсивных функций. Прокомментируем более подробно шаги 2) и 3).

Функцию $f(\vec{x}\,)$ называем *представимой в* T, если для некоторой формулы $\phi_f(\vec{x}, y)$ выполнено: для любого $\vec{n}$, если $m = f(\vec{n}\,)$, то

$$T \vdash \forall y\, \big(\varphi_f(\underline{\vec{n}}, y) \leftrightarrow y = \underline{m}\,\big).$$

Функция f называется Σ_1-*представимой*, если φ_f может быть выбрана в классе Σ_1-формул. Имеет место следующая лемма.

ЛЕММА 8. *Всякая вычислимая функция* f Σ_1-*представима в* R.

Эта лемма обычно доказывается непосредственно, индукцией по построению частично рекурсивной схемы, определяющей f. Ее можно также легко вывести и из следствия 2, модифицируя Σ_1-определение графика функции f. Если формула $\exists z\, \varphi_0(\vec{x}, y, z)$ представляет график f, где $\varphi_0 \in \Delta_0$, то в качестве φ_f достаточно взять формулу, выражающую тот факт, что y есть первый элемент минимальной пары $\langle y, z\rangle$, для которой $\varphi_0(\vec{x}, y, z)$.

Заметим, что представимость характеристической функции множества A в T влечет его разрешимость в T, поэтому лемма 8 влечет разрешимость в R всех алгоритмически разрешимых множеств. Рассматривая частичные $\{0,1\}$-значные функции мы аналогично можем вывести из леммы 8 представимость в R всех перечислимых множеств и отделимость пар перечислимых множеств.

Лемма о неподвижной точке. Центральным элементом доказательства Гёделя было построение предложения ψ, утверждающего собственную недоказуемость. Возможность построения арифметических предложений, ссылающихся на самих себя, не зависит от специфики формулы доказуемости и фактически имеет место для любого свойства, выразимого формулой рассматриваемого языка. Соответствующее утверждение в современных изложениях выделяют в отдельную *лемму о неподвижной точке.*

Пусть T – арифметическая теория, а $\varphi(x)$ – формула языка T с числовой переменной x, возможно, содержащая и другие свободные переменные. Для любой формулы ψ языка T обозначим через $\varphi[\psi]$ формулу $\varphi(\underline{m}\,)$, где $m = \ulcorner\psi\urcorner$ – гёделев номер ψ. Как всегда, мы предполагаем, что гёделева нумерация формул вычислима.

Лемма 9. *Пусть T – арифметическая теория, содержащая* R. *Тогда для любой формулы $\varphi(x)$ языка T найдется формула ψ с теми же свободными переменными, что и φ, кроме x, такая, что*

$$T \vdash \psi \leftrightarrow \varphi[\psi].$$

Доказательство. Рассмотрим функцию, сопоставляющую формуле $\theta(x)$ формулу $\theta[\theta]$. Эта функция вычислима, поскольку вычислима гёделева нумерация формул. Следовательно, будет вычислимой и функция

$$\ulcorner\theta\urcorner \mapsto \ulcorner\theta[\theta]\urcorner.$$

Пусть $\mathrm{Diag}(x,y)$ обозначает Σ_1-формулу, представляющую эту функцию в R. Рассмотрим формулу

$$\varphi'(x) = \exists y\,\bigl(\varphi(y) \wedge \mathrm{Diag}(x,y)\bigr),$$

и пусть $\psi = \varphi'[\varphi']$. По определению Diag мы имеем:

$$\mathsf{Q} \vdash \forall y\,\bigl(\mathrm{Diag}(\underline{\ulcorner\varphi'\urcorner}, y) \leftrightarrow y = \underline{\ulcorner\psi\urcorner}\bigr).$$

Отсюда по правилам логики предикатов с равенством получаем, что формула $\varphi'[\varphi']$ эквивалентна $\exists y\bigl(\varphi(y) \wedge y = \underline{\ulcorner\psi\urcorner}\bigr)$, т. е. $\varphi[\psi]$, что и требовалось доказать. Лемма доказана.

Анализ доказательства Гёделя. Рассмотрим доказательство Гёделя с вычислительной точки зрения. Доказуемость в Σ_1-корректной перечислимой теории T, содержащей R, доставляет еще одну универсальную модель вычислимости (следствие 2). Поэтому перечислимые множества можно было бы *определить* как множества, нумеруемые в T. В соответствии с этой точкой зрения *программой*, принимающей множество A, мы можем считать формулу $\varphi_A(x)$, нумерующую A в T:

$$n \in A \iff T \vdash \varphi_A(\underline{n}\,).$$

Гёделев номер любой такой формулы φ_A считается индексом множества A. Вычислением программы φ_A на входе n в этом случае является поиск вывода формулы $\varphi_A(\underline{n}\,)$ в T.

Для такой вычислительной модели можно переформулировать все общие результаты, известные в теории вычислимости, для любой стандартной модели, например для машин Тьюринга. Вспомним, что креативное множество K определялось как $\{m \in \mathbb{N} \mid m \in W_m\}$. В качестве W_m теперь фигурирует множество

$$\{n \in \mathbb{N} \mid T \vdash \varphi_A(\underline{n}\,)\},$$

где $\ulcorner\varphi_A\urcorner = m$, что равносильно

$$\bigl\{n \in \mathbb{N} \mid \mathbb{N} \vDash \exists z\,\bigl(\mathrm{Pr}_T(z) \wedge \mathrm{Sub}(m,n,z)\bigr)\bigr\},$$

где $\mathrm{Sub}(x,y,z)$ представляет функцию $\ulcorner\varphi\urcorner,\ n \mapsto \ulcorner\varphi(\underline{n}\,)\urcorner$. Соответственно, аналог множества K нумеруется формулой

$$\exists z\,\bigl(\mathrm{Pr}_T(z) \wedge \mathrm{Sub}(x,x,z)\bigr). \qquad (7)$$

Повторяя эффективную версию доказательства теоремы 7, мы рассматриваем перечислимое множество, нумеруемое отрицанием формулы (7), и вычисляем его индекс (гёделев номер) m. В соответствии с определением K предложение

$$\neg\exists z\,\big(\mathrm{Pr}_T(z)\wedge \mathrm{Sub}(\underline{m},\underline{m},z)\big)$$

будет независимым. Внимательный читатель уже наверняка обнаружил, что формула $\mathrm{Sub}(x,x,y)$ есть не что иное, как $\mathrm{Diag}(x,y)$, и мы только что буквально выписали решение уравнения на неподвижную точку

$$T\vdash\psi\leftrightarrow\neg\mathrm{Pr}_T(\ulcorner\psi\urcorner),$$

данное в доказательстве леммы 9. Таким образом, гёделевское доказательство теоремы о неполноте сводится, по существу, к эффективной версии теоретико-рекурсивного доказательства. Единственное отличие заключается в выборе Гёделем специфической модели вычислимости.

Приведенный анализ также проливает некоторый свет на роль понятия представимой функции. Мы могли бы обойтись без этого понятия и для доказательства теоремы 7 гёделевским методом, ослабив формулировку леммы о неподвижной точке. Если T является Σ_1-корректной, то нам достаточно истинности эквивалентности $\psi\leftrightarrow\neg\mathrm{Pr}_T(\ulcorner\psi\urcorner)$ в стандартной модели арифметики, а это мы получаем уже из простой нумеруемости отношения Sub некоторой Σ_1-формулой. Другое дело, что более сильная формулировка позволяет утверждать недоказуемость формулы ψ уже при условии простой непротиворечивости T, а это принципиально, например, при доказательстве второй теоремы Гёделя.

Отметим, что первоначальное доказательство теоремы Гёделя–Россера [18] также основывается на применении леммы о неподвижной точке. Россер рассмотрел следующую *россеровскую формулу доказательств* в теории T:

$$\mathrm{Prf}^R_T(x,y)\quad\stackrel{\mathrm{def}}{\iff}\quad\mathrm{Prf}_T(x,y)\wedge\forall z<y\,\forall u<z\,\big(\mathrm{Neg}(x,u)\,\rightarrow\,\neg\mathrm{Prf}_T(u,z)\big),$$

где $\mathrm{Neg}(x,u)$ представляет функцию навешивания отрицания на формулу x: $\ulcorner\varphi\urcorner\mapsto\ulcorner\neg\varphi\urcorner$. Заметим, что если теория T непротиворечива, то $\mathrm{Prf}^R_T(x,y)$ выражает в стандартной модели тот же самый предикат, что и гёделевская формула $\mathrm{Prf}_T(x,y)$. Россеровская формула доказуемости определяется как $\mathrm{Pr}^R_T(x)\stackrel{\mathrm{def}}{\iff}\exists y\,\mathrm{Prf}^R_T(x,y)$.

Вслед за Россером нетрудно показать, что если

$$T\vdash\theta\leftrightarrow\neg\mathrm{Pr}^R_T[\theta], \tag{8}$$

то формула θ не доказуема и не опровержима в T уже при условии простой непротиворечивости T.

Это доказательство также сводится к теоретико-рекурсивному при выборе подходящей модели вычислимости. В данном случае мы работаем с понятием вычислимой $\{0,1\}$-значной частичной функции и рассматриваем формулу $\varphi(x)$

как программу для вычисления функции

$$f(n) = \begin{cases} 0, & \text{если } T \vdash \varphi(\underline{n}\,), \\ 1, & \text{если } T \vdash \neg\varphi(\underline{n}\,), \\ \text{не определено} & \text{в остальных случаях.} \end{cases}$$

Анализ, аналогичный приведенному выше, показывает, что независимое утверждение, получаемое из эффективной версии теоремы Россера, буквально соответствует решению θ уравнения на неподвижную точку (8).

7. Теорема о неполноте и алгоритмические проблемы

Пользуясь изложенной общей схемой доказательства теоремы Гёделя, можно конвертировать в теоремы о неполноте любые из известных неразрешимых алгоритмических проблем в математике. В качестве типичной иллюстрации рассмотрим несколько примеров из арифметики и теории групп.

Десятая проблема Гильберта: по данному диофантову уравнению определить, существует ли у него хотя бы одно целочисленное решение. Этот пример относится к арифметике и потому наиболее близок к традиционной версии теоремы Гёделя. Он интересен тем, что дает самую простую в логическом смысле форму независимого утверждения (для стандартного языка арифметики).

В силу теоремы Матиясевича множество K выразимо некоторой диофантовой формулой вида $\exists \vec{y}\, p(x, \vec{y}\,) = q(x, \vec{y}\,)$, где p, q – многочлены с натуральными коэффициентами. Отрицание этой формулы определяет неперечислимое множество, откуда для данной перечислимой непротиворечивой арифметической теории T мы получаем константу c, для которой утверждение $\forall \vec{x}\; p(\underline{c}, \vec{y}\,) \neq q(\underline{c}, \vec{y}\,)$ истинно, но не доказуемо в T. В работе [53] на основе предшествующих результатов Дж. Джонса была выписана явно вполне обозримая формула (легко конвертируемая в диофантову), для которой это верно. Константа c также может быть указана явно, если задана машина Тьюринга, перечисляющая T, и фиксирована арифметизация машин Тьюринга.

Проблема равенства слов в группах. Пусть фиксировано задание некоторой группы G конечной системой порождающих элементов и определяющих соотношений. Для данного слова w в групповом алфавите

$$\{a_1, a_1^{-1}, \ldots, a_n, a_n^{-1}\},$$

состоящем из порождающих и обратных к ним, требуется определить, выполняется ли равенство $w = 1$ в группе G. Классический результат П. С. Новикова [54] состоит в построении конечно заданной группы G, для которой эта алгоритмическая проблема неразрешима.

С помощью стандартной гёделевой нумерации слов в алфавите группы G мы можем определить в языке арифметики перечислимый неразрешимый предикат $w =_G 1$, выражающий равенство слова w единице в G. В соответствии с общей схемой для заданной (перечислимой непротиворечивой) теории T мы получаем конкретное слово $w_T \neq_G 1$, для которого этот факт не доказуем в T.

Нераспознаваемость инвариантных свойств группы по ее конечному заданию. В силу известной теоремы Адяна–Рабина (см. [55], [56]), почти все нетривиальные инвариантные свойства групп не распознаваемы по их конечным заданиям. Например, таковым является свойство группы быть изоморфной единичной группе.

Рассмотрим естественную гёделеву нумерацию всех конечных заданий групп порождающими элементами и определяющими соотношениями. Обобщая предыдущий пример, мы можем выразить в языке арифметики перечислимый двухместный предикат $w =_G 1$, где G – конечное задание группы, а w – слово в алфавите G. Заметим, что свойство G задавать единичную группу перечислимо, поскольку равносильно утверждению $\bigwedge_{i=1}^{n}(a_i =_G 1)$, где $a_1, \dots, a_n$ – порождающие группы G. Тем самым, свойство *неединичности* группы неперечислимо и выразимо арифметической Π_1-формулой. Общая схема доказательства теоремы Гёделя для любой перечислимой теории T дает пример конечно заданной неединичной группы G_T, неединичность которой не доказуема в T при условии непротиворечивости T.

Аналогично строятся примеры конечно заданных бесконечных групп, бесконечность которых не доказуема в T; групп без кручения, для которых этот факт не доказуем в T, и так далее. Список подобных примеров можно продолжать. Так же как и теорема Чейтина, они мало добавляют нового к нашему пониманию доказуемости в формальных системах. С другой стороны, они демонстрируют наличие недоказуемых фактов, касающихся самых разных областей математики, в том числе и весьма далеких от арифметики.

Типичное недоказуемое утверждение A имеет вид “данный объект C обладает свойством P”, при этом свойство P может быть математически очень естественным. Однако конкретные объекты C, для которых A недоказуемо, как правило, довольно велики, поскольку в основе практически всех известных примеров неперечислимых множеств в конечном итоге лежит кодирование программ (для некоторой универсальной модели вычислений). Получающиеся примеры недоказуемых утверждений, соответственно, зависят от довольно больших констант. В свою очередь эти константы зависят от конкретной гёделевой нумерации программ – например, машин Тьюринга – и потому не являются ни в каком смысле “каноническими”. Это несколько снижает ценность таких примеров с точки зрения обычной математики.

Интересен вопрос о том, можно ли все-таки использовать эти константы для какой-либо содержательной классификации формальных систем, например, зафиксировав какую-то конкретную “естественную” гёделеву нумерацию. Помимо работ Чейтина высказывания на эту тему можно встретить и у других авторов, например в [53]. Работающие примеры классификаций такого рода автору данного обзора не известны, однако принципиально этот вопрос пока можно считать открытым.[12]

[12] Примеры того, что при некотором (неестественном) выборе гёделевой нумерации в теореме Чейтина все константы можно обратить в нуль, показывают, что попытки подобных классификаций не могут быть совсем наивными. Однако это еще не исключает возможных более сложных решений задачи.

8. Математически естественные примеры недоказуемых утверждений

Несколько по-иному обстоит дело с естественными примерами недоказуемых утверждений, которые были найдены в тех или иных областях математики на протяжении всего периода после открытия Гёделем своих теорем.

Самым знаменитым примером утверждения такого рода была континуум-гипотеза Кантора. Непротиворечивость этой гипотезы аксиомам теории множеств ZFC была доказана Гёделем, а ее недоказуемость в ZFC – Полом Коэном (оба результата при условии непротиворечивости самой ZFC). Впоследствии была установлена независимость от аксиом ZFC многих других утверждений в теории множеств, общей топологии и теории функций. Все эти результаты были получены на основе других, более глубоких методов, связанных со спецификой теории множеств. В частности, естественные независимые утверждения в теории множеств существенным образом связаны с бесконечными множествами, что принципиально отличает их от гёделевских независимых формул, имеющих финитный [13] (и выразимый в языке арифметики) характер.

Естественные математические утверждения финитного – как правило, комбинаторного – характера, независимые от стандартных логических теорий, были обнаружены значительно позже. В качестве теорий при этом рассматривались арифметика Пеано PA, предикативная арифметика второго порядка ATR_0 или даже теория множеств ZFC вместе с некоторыми аксиомами больших кардиналов. Каждый пример финитного независимого утверждения основывается на глубоком анализе конкретной рассматриваемой формальной системы и не обобщается на произвольные перечислимые теории. В этом смысле здесь речь идет о результатах другого, *негёделевского типа*.

Отметим, что из таких результатов вытекает неполнота ряда важных конкретных теорий, например PA или даже ZFC, однако невозможно вывести факт их *принципиальной непополняемости*, т. е. первую теорему Гёделя в той степени общности, которую дал сам Гёдель. Тем не менее, такая потеря в общности с лихвой компенсируется естественностью утверждений, независимость которых устанавливается. Ниже мы опишем лишь некоторые наиболее известные примеры такого рода, однако сколько-нибудь подробный обзор этой темы выходит далеко за рамки данной статьи (в качестве доступного введения в предмет можно рекомендовать, например, [57]).

Принцип Париса–Харрингтона. Одним из первых и наиболее впечатляющих примеров математически естественного финитного независимого утверждения был так называемый *принцип Париса–Харрингтона*, обобщающий конечную теорему Рамсея и найденный в работе [58].

Обозначим через $[X]^k$ множество всех k-элементных подмножеств конечного множества X, а через $|X|$ его мощность. В качестве X мы будем рассматривать

[13] т. е. формулируемый без использования бесконечных объектов; мы употребляем здесь слово "финитный" в более слабом смысле, чем это принято в рамках обсуждения программы Гильберта.

начальный отрезок натурального ряда $\{0, 1, \ldots, n-1\}$, обозначаемый $\mathbf{n}$. Рассматриваются раскраски множества $[X]^k$ в c цветов, т. е. функции $f\colon [X]^k \to \mathbf{c}$. Подмножество $Y \subseteq X$ называется *f-одноцветным*, если ограничение функции f на подмножество $[Y]^k \subseteq [X]^k$ постоянно.

Классическая теорема Рамсея утверждает, что для каждых c, k, m найдется такое n, что для любой раскраски $f\colon [\mathbf{n}]^k \to \mathbf{c}$ найдется одноцветное $Y \subseteq \mathbf{n}$, для которого $|Y| \geqslant m$. Хорошо известно, что эту теорему можно формализовать и доказать в арифметике Пеано PA или даже в более слабой примитивно рекурсивной арифметике.

Принцип Париса–Харрингтона PH отличается от этой теоремы лишь дополнительным требованием на одноцветное множество $Y \subseteq \mathbf{n}$ в ее заключении: помимо $|Y| \geqslant m$ должно быть выполнено $|Y| > \min(Y)$. Так же как и классическая теорема, этот принцип легко выводится из бесконечной теоремы Рамсея на основе соображений компактности. Тем более удивительно, что, как показали Парис и Харрингтон, принцип PH уже не доказуем в арифметике Пеано.

Помимо принципа Париса–Харрингтона, для арифметики Пеано известно еще несколько характерных невыводимых истинных комбинаторных утверждений: “последовательность Гудстейна”, “игра ‘Геракл и Гидра’” [59], “принцип Канамори–Макалуна” [60], “принцип Червя” [61], [62] и другие. Все эти принципы с логической точки зрения равносильны утверждению об 1-непротиворечивости арифметики PA и потому доказуемы в более сильных теориях, таких как ATR_0 и ZFC.

Финитный вариант теоремы Крускала. Наиболее известный пример финитного комбинаторного утверждения, независимого от теории ATR_0, связан с теоремой Крускала о деревьях. Это утверждение было найдено Харви Фридманом и опубликовано в работе [63].

Будем рассматривать конечные деревья $(T, \leqslant, \inf)$ как частично упорядоченные множества с операцией inf, сопоставляющей элементам $x, y \in T$ их точную нижнюю грань $\inf(x, y)$. Назовем *гомеоморфным вложением* дерева T_1 в дерево T_2 инъективное отображение $f\colon T_1 \to T_2$, сохраняющее inf (и тем самым отношение порядка): $f\bigl(\inf(x, y)\bigr) = \inf\bigl(f(x), f(y)\bigr)$.

Теорема Крускала, в простейшей форме, состоит в том, что для любой бесконечной последовательности конечных деревьев $T_0, T_1, \ldots$ найдутся такие индексы $i < j$, что T_i гомеоморфно вложимо в T_j. Отметим, что это утверждение не является финитным, так как апеллирует к произвольным бесконечным последовательностям (конечных деревьев).

Рассмотрим следующее финитное следствие теоремы Крускала. Для любого m найдется n такое, что если $T_0, T_1, \ldots, T_n$ – последовательность конечных деревьев, в которой каждое дерево T_k имеет не более $m + k$ вершин, то T_i гомеоморфно вложимо в T_j для некоторых $i < j \leqslant n$. Это утверждение легко вытекает из теоремы Крускала по соображениям компактности. С другой стороны, как показал Фридман, оно влечет 1-непротиворечивость ATR_0 и потому не доказуемо в ATR_0 в силу второй теоремы Гёделя. Фактически, финитная форма теоремы Крускала выходит достаточно далеко за рамки теории ATR_0.

В работах [64] и [65] получена точная характеризация бесконечной и конечной теоремы Крускала в терминах теоретико-доказательственных ординалов и быстрорастущих функций. (Точную границу дает так называемый *малый ординал Веблена* и функция расширенной иерархии Гжегорчика с соответствующим индексом.)

Фридман также нашел более сильные финитные комбинаторные утверждения, аналогичные теореме Крускала, связанные со специфическим понятием вложения (так называемым *гэп-вложением*) для помеченных конечных деревьев [63]. Известны и другие аналогичные по силе интересные комбинаторные принципы, например, так называемая "гидра Бухгольца", обобщающая игру "Геракл и Гидра" на помеченные деревья [66].

Финитные утверждения, независимые от ZFC. В последующие годы Фридманом был получен целый ряд новых финитных утверждений, независимых от все более сильных теорий, в том числе и для ZFC вместе с некоторыми аксиомами больших кардиналов. (Эти утверждения доказуемы в предположении других, более сильных, аксиом больших кардиналов.) Часть этих результатов опубликована в [67], однако большая часть доступна лишь в виде предварительных сообщений (см. [68]). Удивительно, что в отличие от традиционных результатов о независимости в теории множеств эти утверждения целиком финитны, однако доказуемы лишь из очень сильных теоретико-множественных предположений, выходящих за рамки ZFC. Более того, в отличие от всех упомянутых выше примеров "математической неполноты", Фридман привел примеры естественных независимых от ZFC утверждений сложности Π_1 в арифметической иерархии. Эти примеры пока также лишь объявлены [68; № 49–51].

9. Приложение: относительные интерпретации

Сначала мы дадим определение интерпретации модели M сигнатуры Ω в другой модели N сигнатуры Σ. Затем мы определим интерпретации между теориями. Наше определение интерпретации довольно общее и допускает параметры, интерпретацию объектов $x \in M$ набором объектов $\vec{x} \in N^k$ и интерпретацию равенства в M отношением конгруэнтности.

Говорим, что задан *перевод* I сигнатуры Ω в сигнатуру Σ, если

1) дана формула $D_I(\vec{x}, \vec{p})$ сигнатуры Σ (наборы $\vec{x}$ и $\vec{p}$ исчерпывают все свободные переменные формулы D_I и могут иметь разную длину k и m соответственно);

2) каждому символу сигнатуры Ω сопоставлена формула сигнатуры Σ соответствующей валентности:

$$\begin{aligned} P &\mapsto P_I(\vec{x}_1, \dots, \vec{x}_n, \vec{p}), \\ f &\mapsto F_I(\vec{x}_1, \dots, \vec{x}_n, \vec{y}, \vec{p}), \\ c &\mapsto C_I(\vec{x}, \vec{p}), \end{aligned}$$

где P, f – n-местные предикатный и функциональный символы, а c – константа из Ω. (Все наборы переменных $\vec{x}_i$ имеют длину k, а $\vec{p}$ – длину m.) В частности, символу равенства в Ω соответствует формула $=_I (\vec{x}, \vec{y}, \vec{p})$ в сигнатуре Σ.

Рассмотрим любую модель $(N; \vec{e})$ сигнатуры Σ с выделенным набором констант $\vec{e}$. Перевод I определяет в N^k множество

$$M_I = \{\vec{a} \in N^k : N \vDash D_I[\vec{a}, \vec{e}]\}$$

вместе с заданными на нем предикатами P_I, F_I и C_I и $=_I$.

Определение 10. Для данного $\vec{e}$ перевод I есть *интерпретация M в N*, если выполнены условия:

1) $=_I (\vec{x}, \vec{y}, \vec{e})$ удовлетворяет в M_I аксиомам равенства для сигнатуры Ω, т. е. определяет отношение конгруэнтности на M_I;

2) предикаты F_I и C_I определяют на множестве M_I по модулю $=_I$ функцию f_I и константу c_I соответственно;

3) модель $(M_I; P_I, f_I, c_I)/ =_I$ изоморфна M.

Набор констант $\vec{e}$, для которого выполнены эти условия, называется *допустимым* для данной интерпретации I. Интерпретация I имеет *определимые параметры*, если для некоторой формулы $\mathrm{Par}_I(\vec{p})$ сигнатуры Σ имеет место:

1) $N \vDash \exists \vec{x}\ \mathrm{Par}_I(\vec{x})$;

2) если $N \vDash \mathrm{Par}_I[\vec{e}]$, то набор $\vec{e}$ допустим для I.

Формулу A будем называть *упрощенной*, если всякая атомарная подформула, входящая в A, имеет вид $P(x_1, \ldots, x_n)$, $f(x_1, \ldots, x_n) = y$ или $c = x$, где $x, x_1, \ldots, x_n, y$ – переменные, P и f – предикатный и функциональный символы сигнатуры, а c – константа. Хорошо известно, что всякая формула в логике первого порядка эквивалентна некоторой упрощенной формуле.

Пусть I – перевод сигнатуры Ω в Σ. Каждой переменной x (языка Ω) сопоставим свой набор переменных $\vec{x}$ (языка Σ) длины k. Определим перевод $A^I(\vec{x}_1, \ldots, \vec{x}_n, \vec{p})$ упрощенной формулы $A(x_1, \ldots, x_n)$ сигнатуры Ω индукцией по построению A:

(i) $P(x_1, \ldots, x_n)^I \overset{\text{def}}{\iff} P_I(\vec{x}_1, \ldots, \vec{x}_n, \vec{p})$,
$(c = x)^I \overset{\text{def}}{\iff} C_I(\vec{x}, \vec{p})$ и $f(x_1, \ldots, x_n) = y \overset{\text{def}}{\iff} F_I(\vec{x}_1, \ldots, \vec{x}_n, \vec{y}, \vec{p})$;

(ii) $(\neg A)^I \overset{\text{def}}{\iff} \neg A^I$, $(A \wedge B)^I \overset{\text{def}}{\iff} (A^I \wedge B^I)$,

(iii) $\big(\forall x\, A(x)\big)^I \overset{\text{def}}{\iff} \forall \vec{x}\, \big(D_I(\vec{x}, \vec{p}) \to A^I(\vec{x}, \vec{p})\big)$,

(iv) $\big(\exists x\, A(x)\big)^I \overset{\text{def}}{\iff} \exists \vec{x}\, \big(D_I(\vec{x}, \vec{p}) \wedge A^I(\vec{x}, \vec{p})\big)$.

Переводом произвольной формулы A сигнатуры Ω считаем перевод эквивалентной ей упрощенной формулы. Эта формула определена однозначно с точностью до логической эквивалентности.

Пусть $I(a)$ означает элемент множества $M_I \subseteq N^k$, соответствующий $a \in M$ при интерпретации I модели M в N. Индукцией по построению A получаем следующее предложение.

Предложение 4. *Для любой формулы A сигнатуры Ω, любых допустимых $\vec{e} \in N$ и любых $a_1, \ldots, a_n \in M$*

$$M \vDash A[a_1, \ldots, a_n] \iff N \vDash A^I[I(a_1), \ldots, I(a_n), \vec{e}].$$

Обозначим через $\mathrm{Th}(N)$ множество всех предложений, истинных в модели M.

Следствие 6. *Если M интерпретируема в N с определимыми параметрами и теория* $\mathrm{Th}(N)$ *разрешима, то такова и* $\mathrm{Th}(M)$.

Доказательство. Для данного предложения A в языке M имеем

$$M \vDash A \iff N \vDash \forall\vec{p}\,\big(\mathrm{Par}_I(\vec{p}) \to A^I(\vec{p})\big).$$

Таким образом, для проверки выполнимости формулы A в M достаточно проверить выполнимость $\forall\vec{p}\,\big(\mathrm{Par}_I(\vec{p}) \to A^I(\vec{p})\big)$ в модели N.

Пусть даны теория T в сигнатуре Ω и U в сигнатуре Σ. Считаем, что T задана множеством аксиом, являющихся замкнутыми формулами.

Определение 11. Перевод I с определимыми параметрами есть *интерпретация теории T в U*, если

1) $U \vdash \mathrm{Par}_I(\vec{p}) \to A^I(\vec{p})$ для любой аксиомы $A \in T$;

2) $U \vdash \mathrm{Par}_I(\vec{p}) \to \exists\vec{x}\, D_I(\vec{x},\vec{p})$;

3) $U \vdash \mathrm{Par}_I(\vec{p}) \to E^I(\vec{p})$, где E – аксиома равенства для сигнатуры Ω;

4) $U \vdash \mathrm{Par}_I(\vec{p}) \to \big(\forall x_1 \ldots x_n\ \exists! y\ f(x_1,\ldots,x_n) = y\big)^I$ для функциональных символов f из Ω;

5) $U \vdash \mathrm{Par}_I(\vec{p}) \to (\exists! x\ c = x)^I$ для констант c из Ω.

Теория T *интерпретируема в U*, если существует интерпретация T в U.

Таким образом, в любой модели N теории U при выборе параметров, удовлетворяющих условию Par_I, перевод I определяет модель теории T.

Индукцией по длине вывода формулы A получаем следующее предложение.

Предложение 5. *Если I – интерпретация T в U и $T \vdash A$, то*

$$U \vdash \forall\vec{x}\,\big(\mathrm{Par}_I(\vec{p}) \wedge D_I(\vec{x},\vec{p}) \;\to\; A^I(\vec{x},\vec{p})\big).$$

Следствие 7. *Если T интерпретируема в U и теория U непротиворечива, то такова и T.*

Этот результат установлен элементарными (синтаксическими) методами, не опирающимися на теорию множеств. Заметим, что доказательства непротиворечивости, опирающиеся на существование модели, как правило, выходят за рамки элементарных методов, так как обычно модели являются бесконечными и строятся в рамках теории множеств. Метод интерпретаций позволяет избавиться от ненужных гипотез о существовании бесконечных множеств и приводит к "финитному" сведению одной теории к другой. Этим способом можно установить, в частности, непротиворечивость элементарной геометрии Лобачевского относительно элементарной евклидовой геометрии, непротиворечивость континуум-гипотезы, так же как и ее отрицания, относительно теории множеств ZFC и другие важные результаты.

Список литературы

[1] K. Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatsh. Math. Phys.*, **38**:1 (1931), 173–198.

[2] R. Zach, "Hilbert's Program", *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta, 2009, http://plato.stanford.edu/archives/spr2009/entries/hilbert-program/.

[3] К. Сморинский, "Теоремы о неполноте", *Справочная книга по математической логике*, т. 4: *Теория доказательств и конструктивная математика*, ред. Дж. Барвайс, Наука, М., 1983, 9–53, 391 с.; пер. с англ.: C. Smorinskii, "The incompleteness theorems", *Handbook of mathematical logic*, Stud. Logic Found. Math., **90**, ed. J. Barwise, North-Holland, Amsterdam–New York–Oxford, 1977, 821–865.

[4] S. C. Kleene, "Introductory note to 1930b, 1931 and 1932b", *Kurt Gödel. Collected Works*, v. 1: *Publications* 1929–1936, ed. S. Feferman et al., The Clarendon Press, Oxford Univ. Press, New York, 1986, 126–141.

[5] K. Gödel, "On undecidable propositions of formal mathematical systems", *Kurt Gödel. Collected Works*, v. 1: *Publications* 1929–1936, The Clarendon Press, Oxford Univ. Press, New York, 1986, 346–373.

[6] *Kurt Gödel. Collected Works*, v. 1: *Publications* 1929–1936, eds. S. Feferman, J. R. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay, J. van Heijenoort, The Clarendon Press, Oxford Univ. Press, New York, 1986.

[7] A. Tarski, "Der Wahrheitsbegriff in der formalisierten Sprachen", *Stud. Philos.*, **1** (1935), 261–405.

[8] A. Ehrenfeucht, S. Feferman, "Representability of recursively enumerable sets in formal theories", *Arch. Math. Logik Grundlag.*, **5**:1-2 (1960), 37–41.

[9] R. M. Smullyan, *Diagonalization and self-reference*, Oxford Logic Guides, **27**, The Clarendon Press, Oxford Univ. Press, New York, 1994, 396 pp.

[10] J. B. Rosser, "Gödel theorems for non-constructive logics", *J. Symbolic Logic*, **2**:3 (1937), 129–137.

[11] W. Craig, "On axiomatizability within a system", *J. Symbolic Logic*, **18**:1 (1953), 30–32.

[12] S. C. Kleene, "General recursive functions of natural numbers", *Math. Ann.*, **112**:1 (1936), 727–742.

[13] S. C. Kleene, "Recursive predicates and quantifiers", *Trans. Amer. Math. Soc.*, **53**:1 (1943), 41–73.

[14] S. C. Kleene, "A symmetric form of Gödel's theorem", *Indag. Math.*, **12** (1950), 244–246.

[15] С. К. Клини, *Введение в метаматематику*, ИЛ, М., 1957, 526 с.; пер. с англ.: S. C. Kleene, *Introduction to metamathematics*, Van Nostrand, New York, 1952, 550 pp.

[16] Р. М. Смальян, *Теория формальных систем*, Математическая логика и основания математики, Наука, М., 1981, 208 с.; пер. с англ.: R. M. Smullyan, *Theory of formal systems*, Ann. Math. Stud., **47**, Princeton Univ. Press, Princeton, NJ, 1961, 142 pp.

[17] В. А. Успенский, "Теорема Гёделя о неполноте в элементарном изложении", *УМН*, **29**:1 (1974), 3–47; англ. пер.: V. A. Uspenskii, "An elementary exposition of Gödel's incompleteness theorem", *Russian Math. Surveys*, **29**:1 (1974), 63–106.

[18] B. Rosser, "Extensions of some theorems of Gödel and Church", *J. Symbolic Logic*, **1**:3 (1936), 87–91.

[19] A. Mostowski, "On definable sets of positive integers", *Fund. Math.*, **34** (1947), 81–112.

[20] W. V. Quine, “Concatenation as a basis for arithmetic”, *J. Symbolic Logic*, **11**:4 (1946), 105–114.

[21] A. Tarski, A. Mostowski, R. M. Robinson, *Undecidable theories*, Stud. Logic Found. Math., North-Holland, Amsterdam, 1953, 98 pp.

[22] A. S. Troelstra, H. Schwichtenberg, *Basic proof theory*, Cambridge Tracts Theoret. Comput. Sci., **43**, Cambridge Univ. Press, Cambridge, 1996, 343 pp.

[23] Дж. Шёнфилд, *Математическая логика*, Наука, М., 1975, 527 с.; пер. с англ.: J. R. Shoenfield, *Mathematical logic*, Addison-Wesley, Reading, MA–London–Don Mills, ON, 1967, 344 pp.

[24] P. Pudlák, “Cuts, consistency statements and interpretations”, *J. Symbolic Logic*, **50**:2 (1985), 423–441.

[25] S. Feferman, “Arithmetization of metamathematics in a general setting”, *Fund. Math.*, **49** (1960), 35–92.

[26] P. Hájek, P. Pudlák, *Metamathematics of first-order arithmetic*, Perspect. Math. Logic, Springer-Verlag, Berlin, 1993, 460 pp.

[27] Ю. В. Матиясевич, “Диофантовость перечислимых множеств”, *Докл. АН СССР*, **191**:2 (1970), 279–282; англ. пер.: Yu. V. Matiyasevich, “Enumerable sets are Diophantine”, *Soviet Math. Dokl.*, **11** (1970), 354–358.

[28] Ю. В. Матиясевич, *Десятая проблема Гильберта*, Наука, М., 1993, 224 с.; англ. пер.: Yu. V. Matiyasevich, *Hilbert's tenth problem*, Found. Comput. Ser., MIT Press, Cambridge, MA, 1993, 264 pp.

[29] Ю. Л. Ершов, *Проблемы разрешимости и конструктивные модели*, Наука, М., 1980, 416 с.

[30] Ю. Л. Ершов, И. А. Лавров, А. Д. Тайманов, М. А. Тайцлин, “Элементарные теории”, *УМН*, **20**:4 (1965), 37–108; англ. пер.: Yu. L. Ershov, I. A. Lavrov, A. D. Taimanov, M. A. Taitslin, “Elementary theories”, *Russian Math. Surveys*, **20**:4 (1965), 35–105.

[31] A. Bès, “A survey of arithmetical definability. A tribute to Maurice Boffa”, *Bull. Belg. Math. Soc. Simon Stevin*, 2001, suppl., 1–54.

[32] V. Švejdar, “An interpretation of Robinson arithmetic in its Grzegorczyk's weaker variant”, *Fund. Inform.*, **81**:1–3 (2007), 347–354.

[33] R. L. Vaught, “On a theorem of Cobham concerning undecidable theories”, *Logic, Methodology and Philosophy of Science* (Proc. 1960 Internat. Congr.), Stanford Univ. Press, Stanford, 1962, 14–25.

[34] V. Švejdar, “Weak theories and essential incompleteness”, *The Logica Yearbook* 2007, Proceedings of the Logica 07 International Conference, ed. M. Peliš, Philosophia, Praha, 2008, 213–224.

[35] A. Visser, “Pairs, sets and sequences in first-order theories”, *Arch. Math. Logic*, **47**:4 (2008), 299–326.

[36] A. Visser, *Cardinal arithmetic in weak theories*, Logic Group Preprint Series, **265**, Department of Philosophy, Univ. Utrecht, 2008, http://igitur-archive.library.uu.nl/lg/2008-0422-200811/UUindex.html.

[37] H. Putnam, “Decidability and essential undecidability”, *J. Symbolic Logic*, **22**:1 (1957), 39–54.

[38] A. Ehrenfeucht, “Two theories with axioms built by means of pleonasms”, *J. Symbolic Logic*, **22**:1 (1957), 36–38.

[39] J. P. Jones, J. C. Shepherdson, “Variants of Robinson's essentially undecidable theory R”, *Arch. Math. Logik Grundlag.*, **23**:1 (1983), 61–64.

[40] A. Visser, *Why the theory R is special?*, Logic Group Preprint Series, **279**, Department of Philosophy, Univ. Utrecht, 2009, http://igitur-archive.library.uu.nl/ph/2009-0812-200111/UUindex.html.

[41] R. L. Vaught, "Axiomatizability by a schema", *J. Symbolic Logic*, **32**:4 (1967), 473–479.

[42] S. Feferman, "Finitary inductively presented logics", *Logic Colloquium'88* (Padova, 1988), Stud. Logic Found. Math., **127**, North-Holland, Amsterdam, 1989, 191–220.

[43] A. Grzegorczyk, "Undecidability without arithmetization", *Studia Logica*, **79**:2 (2005), 163–230.

[44] F. Ferreira, "Polynomial time computable arithmetic", *Logic and computation* (Pittsburgh, PA, 1987), Contemp. Math., **106**, Amer. Math. Soc., Providence, RI, 1990, 137–156.

[45] Ю. Л. Ершов, *Определимость и вычислимость*, Сибирская школа алгебры и логики, Научная книга, Новосибирск, 1996, 287 с.; англ. пер.: Yu. L. Ershov, *Definability and computability*, Siberian School of Algebra and Logic, Consultants Bureau, New York, 1996, 262 pp.

[46] G. J. Chaitin, "Information-theoretic limitations of formal systems", *J. Assoc. Comput. Mach.*, **21**:3 (1974), 403–424.

[47] G. Boolos, "A new proof of Gödel's incompleteness theorem", *Notices Amer. Math. Soc.*, **36** (1989), 388–390; G. Boolos, "A new proof of Gödel's incompleteness theorem", *Logic, logic, and logic*, Harvard Univ. Press, Cambridge, MA, 1998, 383–388.

[48] P. Raatikainen, "On interpreting Chaitin's incompleteness theorem", *J. Philos. Logic*, **27**:6 (1998), 569–586.

[49] M. Kikuchi, "Kolmogorov complexity and the second incompleteness theorem", *Arch. Math. Logic*, **36**:6 (1997), 437–443.

[50] В. А. Успенский, "Теорема Гёделя о неполноте и четыре дороги, ведущие к ней. Лекция 1", *Лекции Летней школы 'Современная математика'*, Дубна, 2007, http://www.mathnet.ru/php/presentation.phtml?option_lang=rus&presentid=122.

[51] В. А. Успенский, *Теорема Гёделя о неполноте*, Популярные лекции по математике, Наука, М., 1982, 110 с.

[52] Э. Мендельсон, *Введение в математическую логику*, Наука, М., 1971, 320 с.; пер. с англ.: E. Mendelson, *Introduction to mathematical logic*, Van Nostrand, Princeton, NJ, 1964, 300 pp.

[53] V. H. Dyson, J. P. Jones, J. C. Shepherdson, "Some Diophantine forms of Gödel's theorem", *Arch. Math. Logik Grundlag.*, **22**:1-2 (1982), 51–60.

[54] П. С. Новиков, "Об алгоритмической неразрешимости проблемы тождества слов в теории групп", Тр. МИАН СССР, **44**, Изд-во АН СССР, М., 1955, 3–143.

[55] С. И. Адян, "Алгоритмическая неразрешимость проблем распознавания некоторых свойств групп", *Докл. АН СССР*, **103**:4 (1955), 533–535.

[56] С. И. Адян, "Неразрешимость некоторых алгоритмических проблем теории групп", Тр. ММО, **6**, 1957, 231–298.

[57] A. Bovykin, "Brief introduction to unprovability", *Logic Colloquium* 2006, Proceedings of Annual European Conference on Logic of the Association for Symbolic Logic (Radboud University, Nijmegen, 2006), Lect. Notes Log., Assoc. Symbol. Logic, Chicago, IL; Cambridge Univ. Press, Cambridge Cambridge Univ. Press, 2009, 38–64.

[58] Дж. Парис, Л. Харрингтон, "Математическая неполнота в арифметике Пеано", *Справочная книга по математической логике*, т. 4: *Теория доказательств и конструктивная математика*, ред. Дж. Барвайс, Наука, М., 1983, 319–327; пер. с англ.: J. Paris, L. Harrington, "A mathematical incompleteness in Peano

arithmetic", *Handbook of mathematical logic*, Stud. Logic Found. Math., **90**, ed. J. Barwise, North-Holland, Amsterdam–New York–Oxford, 1977, 1133–1142.

[59] L. A. S. Kirby, J. Paris, "Accessible independence results for Peano arithmetic", *Bull. London Math. Soc.*, **14**:4 (1982), 285–293.

[60] A. Kanamori, K. McAloon, "On Gödel's incompleteness and finite combinatorics", *Ann. Pure Appl. Logic*, **33**:1 (1987), 23–41.

[61] L. D. Beklemishev, "The Worm principle", *Logic Colloquium'*02, Joint proceedings of the Annual European Summer Meeting of the Association for Symbolic Logic and the Biannual Meeting of the German Association for Mathematical Logic and the Foundations of Exact Sciences (the Colloquium Logicum) (Munster, 2002), Lect. Notes Log., **27**, Assoc. Symbol. Logic, La Jolla, CA; A K Peters, Ltd., Wellesley, MA, 2006, 75–95; Logic Group Preprint Series, **219**, Utrecht Univ., March, 2003.

[62] M. Hamano, M. Okada, "A relationship among Gentzen's proof-reduction, Kirby–Paris' hydra game, and Buchholz's hydra game", *Math. Logic Quart.*, **43**:1 (1997), 103–120.

[63] S. G. Simpson, "Nichtbeweisbarkeit von gewissen kombintorischen Eigenschaften endlicher Bäume", *Arch. Math. Logik Grundlag.*, **25**:1 (1985), 45–65.

[64] M. Rathjen, A. Weiermann, "Proof-theoretic investigations on Kruskal's theorem", *Ann. Pure Appl. Logic*, **60**:1 (1993), 49–88.

[65] H. M. Friedman, "Internal finite tree embeddings", *Reflections on the foundations of mathematics* (Stanford, CA, 1998), Lect. Notes Log., **15**, Assoc. Symbol. Logic, Urbana, IL; A K Peters, Ltd., Natick, MA, 2002, 60–91.

[66] W. Buchholz, "An independence result for (Π^1_1-CA) + BI", *Ann. Pure Appl. Logic*, **33**:2 (1987), 131–155.

[67] H. M. Friedman, "Finite functions and the necessary use of large cardinals", *Ann. of Math.* (2), **148**:3 (1998), 803–893, arXiv: arXiv: math/9811187.

[68] H. M. Friedman, Downloadable manuscripts at `http://www.math.ohio-state.edu/~friedman/manuscripts.html`.

Л. Д. Беклемишев (L. D. Beklemishev)
Математический институт им. В. А. Стеклова РАН
E-mail: `bekl@mi.ras.ru`

Поступила в редакцию
20.08.2010