VII Летняя школа «Современная математика»

# Владимир Андреевич Успенский

# Теорема Гёделя о неполноте и четыре дороги, ведущие к ней

Ратмино, 21 июля 2007 года

## Дорога ГЁДЕЛЯ (проложена в 1930 г.)

Эта дорога основана на парадоксе лжеца, называемого также парадоксом Эпименида. Апостол Павел в послании к Титу (1:12), имея в виду критян, говорит: “Из них же самих один $\langle\ldots\rangle$ сказал: ”Критяне всегда лжецы $\langle\ldots\rangle$””. Есть версия, что сказавшего звали Эпименид. Если считать его слова верными, то получается, что он и сам лжец. Здесь парадокс ещё не возникает. Парадокс возникнет, если предположить, что кроме него критян больше не было, и единственный критянин ничего другого не сказал. Примерно в такой форме парадокс возник в IV веке до нашей эры у древнегреческого софиста Эвбулида, которому приписывают такое высказывание: “Высказывание, которое я сейчас произношу, ложно”. Попытка выяснить, истинно оно или ложно, приводит к противоречию; поэтому, с точки зрения логики, оно вообще не есть высказывание. Основываясь на этой идее, Гёдель нашёл формулу, выражающую недоказуемость самой себя. Легко понять, что эта формула истинна, но не доказуема. В самом деле, если бы она была доказуема, то выражала бы истину и, следовательно, была бы недоказуема. Значит, она недоказуема. Значит, формула, выражающая эту недоказуемость (то есть она сама) истинна. Вот мы и нашли истинную, но недоказуемую формулу, существование которой провозглашает теорема Гёделя.

Посмотрим, как Гёдель построил такую формулу.

Напомним, что любое перечислимое множество кортежей натуральных чисел фиксисрованной длины выражается некоторой формулой рассматриваемого языка. Поскольку выразимость функции означает, по определению, выразимость её графика, то и каждая вычислимая функция $k$ натуральных переменных выражается подходящей формулой языка.

Пусть $A$ — формула. Через обозначаем $A(r)$ обозначаем результат подстановки в формулу $A$ числа $r$ вместо параметра $x$, через $A(r,s)$ — результат подстановки в $A$ чисел $r$ и $s$ вместо параметров $x$ и $y$, через $A(r,s,t)$ — результат подстановки в $A$ чисел $r$, $s$ и $t$ вместо вместо параметров $x$, $y$ и $z$. Ясно, что говоря о подстановке чисел, мы подставляем не сами числа, а их обозначения (цифры) в виде нолей со штрихами.

1. Нумеруем все формулы:

$$C_1,\ C_2,\ C_3,\ \ldots$$

2. Нумеруем все формальные доказательства:

$$\Gamma_1,\ \Gamma_2,\ \Gamma_3,\ \ldots$$

3. Рассматриваем функцию $g$:

$$g(s) = t$$

тогда и только тогда, когда

$\Gamma_s$ есть доказательство формулы $C_t$.

Ясно, что функция $g$ вычислима. Поэтому её выражает некоторая формула $G$ с параметрами $x$ и $y$. Когда, для каких пар чисел формула $G(s,t)$ истинна? Она истинна тогда только тогда, когда доказательство с номером $s$ является доказательством формулы с номером $t$.

4. Для каждого числа $t$ формула $\exists w\ G(w,t)$ означает доказуемость формулы $C_t$.

5. Для каждого числа $t$ формула $\neg\exists w\ G(w,t)$ означает недоказуемость формулы $C_t$.

6. Нумеруем все открытые формулы с единственным параметром $x$:

$$A_1,\ A_2,\ A_3,\ \ldots$$

7. Рассматриваем функцию $f$:
$f(m,n)$ есть номер формулы $A_m(n)$. Ясно, что $f$ — вычислимая функция. Поэтому её выражает некоторая формула $F$ с параметрами $x$, $y$ и $z$. Когда, для каких троек чисел формула $F(m,n,p)$ истинна? Она истинна тогда и только тогда, когда $p$ есть номер формулы $A_m(n)$.

8. Для чисел $e$, $m$, $n$ формула

$$\exists u\ [G(e,u) \& F(m,n,u)]$$

означает, что число $e$ есть номер доказательства формулы $A_m(n)$, имеющей номер $f(m,n)$.

9. Для чисел $m$, $n$ формула

$$\neg\exists w\ \exists u\ [G(w,u) \& F(m,n,u)]$$

означает недоказуемость формулы $A_m(n)$, имеющей номер $f(m,n)$.

10. Для каждого числа $n$ формула

$$\neg\exists w\ \exists u\ [G(w,u) \& F(n,n,u)]$$

означает недоказуемость формулы $A_n(n)$, имеющей номер $f(n,n)$.

11. Теперь рассматриваем формулу

$$\neg\exists w\ \exists u\ [G(w,u)\&F(x,x,u)].$$

с единственным параметром $x$. Эта формула есть $A_q$ при некотором $q$.

12. В силу п. 7 формула $A_q(q)$, то есть формула

$$\neg\exists w\ \exists u\ [G(w,u)\&F(q,q,u)],$$

имеет номер $f(q,q)$.

13. В силу п. 10 формула

$$\neg\exists w\ \exists u\ [G(w,u)\&F(q,q,u)],$$

означает недоказуемость формулы с номером $f(q,q)$.

14. Сравнивая пп. 12 и 13, убеждаемся, что присутствующая в них формула (одна и та же!) означает недоказуемость самой себя. Построение закончено.

## Дорога КОЛМОГОРОВА (указана в 1952 г.)

Эта дорога имеет своим фундаментом некоторые главные теоремы теории алгоритмов. Это нижеследующие теоремы 1, 2, 3 и 4.

Т е о р е м а 1. Всякое разрешимое множество перечислимо.

Т е о р е м а 2. Всякое бесконечное перечислимое множество можно расположить в вычислимую последовательность без повторяющихся членов, то есть снабдить его взаимно однозначной вычислимой в обе стороны нумерацией.

Мы уже пользовались этими теоремами, идя по первой дороге, — когда нумеровали 1) формулы, 2) доказательства, 3) формулы с одним параметром.

Т е о р е м а 3. Образ и полный прообраз перечислимого множества при вычислимом отображении (то есть при отображении, осуществляемым вычислимой функцией) являются перечислимыми множествами.

Для функции $f$:

образ множества $A$ есть множество

$$f(A) \;=\; \{y \mid \exists x \in A\ f(x) = y\};$$

полный прообраз множества $A$ есть множество

$$f^{-1}(A) \;=\; \{x \mid \exists y \in A\ f(x) = y\}.$$

Здесь не требуется, чтобы функция $f$ была всюду определена, так что отображение может быть и частичным.

Ф у н д а м е н т а л ь н о е с л е д с т в и е. **Множество всех доказуемых формул перечислимо.**

Это следствие играет решающую роль на второй, третьей и четвёртой дорогах.

Д о к а з а т е л ь с т в о. Множество всех доказуемых формул есть $\delta(D)$, где $D$ — множество всех доказательств, а $\delta$ — функция выделения доказанной формулы. Множество $D$ разрешимо, а значит, по Теореме 1 перечислимо. Функция $\delta$ вычислима. Применяем Теорему 3.

Т е о р е м а 4. Существует перечислимое множество натуральных чисел с неперечислимым дополнением (до всего натурального ряда).

С л е д с т в и е. Существует выразимое неперечислимое множество натуральных чисел.

Д о к а з а т е л ь с т в о. Оно выражается формулой, служащей отрицанием формулы, выражающей перечислимое множество из Теоремы 4.

С л е д с т в и е  с л е д с т в и я. Множество всех истинных формул неперечислимо.

Д о к а з а т е л ь с т в о. Обозначим через $K$ выразимое неперечислимое множество $K \subset \mathbf{N}$, а через $B$ ту формулу с параметром $x$, которая его выражает. Обозначим через $T$ множество всех истинных формул языка. Обозначим через $f$ вычислимую функцию, относящую каждому числу $n \in \mathbf{N}$ формулу $B(n)$ (напомним, что это есть результат подстановки в формулу $B$ числа $n$ вместо параметра $x$). Очевидно, что

$$K \ = \ f^{-1}(T).$$

Если бы $T$ было перечислимым, то по теореме 3 перечислимым было бы и $K$.

Сопоставляя Фундаментальное следствие и Следствие из следствия, получаем теорему Гёделя.

## Дорога ЧЕЙТИНА (проложена в 1974 г.)

Эта дорога основана на парадоксе, предложенном библиотекарем оксфордского университета Бéрри в самом начале XX века. Парадокс Берри возникает из следующего описания некоторого натурального числа: “Наименьшее натуральное число, которое нельзя описать посредством менее ста слов”. Попытка выяснить, допускает или не допускает само это число описания посредством менее ста слов, приводит к противоречию. Для того, чтобы из этого парадокса возникла теорема Гёделя, надо воспользоваться понятием *колмогоровской сложности* слова. Понятие колмогоровской сложности ввели независимо Соломонофф, Колмогоров и Чейтин в начале 60-х годов. В предположении, что теорема Гёделя неверна, оказывается возможным предъявить число с парадоксальными свойствами: оно имеет сложность гораздо меньшую, чем та, наличие которой у него можно доказать. Это и есть сложностная версия парадокса Берри. (Мы исходим из презумпции, что всё доказумое истинно, поэтому парадоксальных чисел не бывает.) Сложностная версия парадокса Берри допускает и такую формулировку, более близкую к нашему изложению: про числа, имеющие очень большую сложность, это их свойство не может быть доказано.

Переходим к математическому воплощению идеи.

Через $\Xi$ обозначается множество всех двоичных слов, то есть $\{0,1\}^*$. Длина слова $\xi$ обозначается так: $|\xi|$. Для каждого слова $\xi \in \Xi$ обозначаем через $\dagger(\xi)$ то натуральное число, для которого $\xi$ служит двоичной записью; если такого числа нет, обозначение не определено. Легко проверить, что

$$|\xi| \ \leqslant\ \log \dagger(\xi) + 1, \tag{1}$$

где логарифм берётся по основанию 2.

Пусть $E$ — произвольное множество, $f$ — произвольная функция (не обязательно всюду определённая) из $\Xi$ в $E$. **Сложность** элемента $e \in E$ **относительно** функции $f$ есть, по определению, число

$$K_f(e) \ = \min\{|\xi| \mid f(\xi) = e\}. \tag{2}$$

Как всегда, минимум по пустому множеству есть $\infty$. (Психологическое пояснение: $f$ рассматривается как функция, дающая объект $e$ по его описанию $\xi$, и сложность — это наименьшая возможная длина описания. Сложность неописуемого объекта бесконечна.)

З а м е ч а н и е. Количество элементов, сложность которых не превосходит заданного числа, конечно.

Скажем, что функция $f$ **не хуже** функции $g$, если с точностью до аддитивной константы, не зависящей от $e$, сложность любого $e$ относительно $f$ не превосходит сложности этого же $e$ относительно $g$:

$$\exists C\ \forall e\ K_f(e) \leqslant K_g(e) + C. \tag{3}$$

Теперь в качестве $E$ берём натуральный ряд $\mathbf{N}$. Тогда можно говорить о вычислимых функциях из $\Xi$ в $E$. ВАЖНО: вычислимые функции не предполагаются всюду определёнными!

О п р е д е л е н и е. Вычислимая функция называется **оптимальной**, если она не хуже любой другой вычислимой функции.

Т е о р е м а. Оптимальные функции существуют.

Фиксируем какую-нибудь оптимальную функцию $f$ и для каждых чисел $e$ и $n$ рассмотрим утверждение

$$K_f(e) > n.$$

В формальном языке это утверждение выражается формулой $\mathbf{I}(n, e)$, получающейся из некоторой формулы $\mathbf{I}$ с двумя параметрами $x$ и $y$ при подстановке вместо этих параметров соответственно чисел $n$ и $e$. Как найти такую формулу $\mathbf{I}$, скажем позже. Пока что мы исходим из того, что такая формула есть. Выпишем её главное свойство:

$$\mathbf{I}(n, e) \ \text{ истинна тогда и только тогда, когда } K_f(e) > n. \tag{$\clubsuit$}$$

Располагаем все доказуемые формулы языка в вычислимую последовательность и фиксируем эту последовательность. Строим вычислимую функцию $g$ из $\Xi$ в $E$ со следующим алгоритмом вычисления. Если $\xi$ не есть двоичная запись натурального числа, оставляем функцию $g$ не определённой на этом $\xi$. Если же $\xi$ есть запись числа $\dagger(\xi)$, то ищем в последовательности доказуемых формул первую формулу вида

$$\mathbf{I}(\dagger(\xi), e) \tag{4}$$

с этим $\xi$ и каким-то $e$. Если и когда мы такую формулу найдём (чего может и не произойти), то берём $e$ в качестве значения функции $g$ на аргументе $\xi$.

Таким образом, все натуральные числа разбиваются на два класса — класс тех чисел, для двоичных записей которых функция $g$ определена, и класс тех чисел, для двоичных записей которых функция $g$ не определена. Число $n$ относится к первому классу в том и только в том случае, если для него существует такое $e$, что формула $\mathbf{I}(n, e)$ доказуема. Наша ближайшая цель — убедиться, что все числа первого класса не превосходят некоторой константы.

Приступаем к построению требуемой оценки. Пусть $n$ принадлежит первому классу. Берём то число $e$ для которого формула (4) доказуема. Раз (4) доказуема, то она истинна. В силу свойства ($\clubsuit$) формулы $\mathbf{I}$ справедливо неравенство

$$K_f(e) \; > \; n. \tag{5}$$

Пусть $n = \dagger(\xi)$. Тогда

$$g(\xi) = e. \tag{6}$$

Оценим сверху сложность числа $e$ относительно функции $g$. Ввиду (6),

$$K_g(e) \; \leqslant \; |\xi|. \tag{7}$$

В сочетании с (1) это даёт

$$K_g(e) \; \leqslant \; \log \dagger(\xi) + 1, \tag{8}$$

или, что то же самое

$$K_g(e) \; \leqslant \; \log n + 1, \tag{9}$$

Оценим сверху сложность того же $e$ относительно функции $f$. Сочетая (3) и (9), получаем:

$$K_f(e) \; \leqslant \; \log n + 1 + C, \tag{10}$$

где $C$ — присутствующая в (3) константа, определяемая сочетанием функций $f$ и $g$. Соединяя (5) и (9), получаем:

$$n \; < \; \log n + 1 + C. \tag{11}$$

Но начиная с некоторого $N$ (зависящего, разумеется, от $C$), неравенство (11) не имеет места. Поэтому все числа первого класса меньше этого $N$.

Теперь мы можем предъявить целую серию истинных, но не доказуемых формул — что и даст теорему Гёделя. Берём произвольное $n$, такое что $n \geqslant N$. В силу сделанного Замечания, существует такое $e$, для которого $K_f(e) \; > \; n$. В силу ($\clubsuit$) будет истинной формула $\mathbf{I}(n, e)$. Однако эта формула будет недоказуемой, так как $n$ — число второго класса. Таким образом, хотя при $n \geqslant N$ числа сложности большей, чем $n$, существуют, ни про одно из них нельзя доказать, что его сложность превосходит $n$.

Осталось выполнить обещание и показать, как строится формула $\mathbf{I}$.

Построение нужной формулы осуществляем в два шага. На первом шагу раскрываем смысл утверждения $K_f(e) > n$ и получаем более подробную его запись. На втором шагу преобразуем эту запись в формулу рассматриваемого языка.

Первый шаг. Неравенство $K_f(e) > n$ равносильно такому утверждению: для всякого $\xi$, для которого $f(\xi) = e$, справедливо неравенство $|\xi| > n$. Записываем это:

$$\forall\xi\ [f(\xi) = e \Rightarrow\ |\xi| > n]. \tag{12}$$

Второй шаг. Запись (12) ещё не является формулой рассматриваемого формального языка. В ней участвуют двоичные слова, а в формулах языка речь может идти только о числах. Легко перейти от слов к числам. Для этого устроим взаимно однозначное и вычислимое соответствие между $\Xi$ и натуральным рядом и вместо того, чтобы говорить о двоичных словах, будем говорить об их номерах, то есть о соотвествующих числах. Соответствие задаётся двумя взаимно обратными вычислимыми функциями: $\pi$ из $\mathbf{N}$ в $\Xi$ и $\rho$ из $\Xi$ в $\mathbf{N}$:

$$\pi(m) = \xi \quad \text{равносильно} \quad \rho(\xi) = m. \tag{13}$$

Вычислимой функции $f$ из $\Xi$ в $\mathbf{N}$ соответствует вычислимая же функция $\hat{f}$ из $\mathbf{N}$ в $\mathbf{N}$, аргументами которой вместо слов служат их номера:

$$\hat{f}(m)\ =\ f(\pi(m)); \qquad f(\xi)\ =\ \hat{f}(\rho(\xi)). \tag{14}$$

График функции $\hat{f}$ является перечислимым множеством, а потому выражается некоторой формулой $F$ с параметрами $x$ и $y$, так что $F(m, e)$ истинна тогда и только тогда, когда $\hat{f}(m) = e$.

Упражнение. Докажите, что (а) множество всех пар $(\xi, n)$, для которых $|\xi| > n$, перечислимо;

(б) множество всех пар $(k, n)$, для которых $|\pi(k)| > n$, перечислимо.

Множество, о котором идёт речь в части (б) Упражнения, в силу своей перечислимости, выражается некоторой формулой $G$ с двумя параметрами. Переходя в записи (12) от двоичных слов к их номерам, переписываем эту запись в виде формулы языка:

$$\forall w\ [F(w, e)\ \Rightarrow\ G(w, n)]. \tag{15}$$

В качестве искомой формулы $\mathbf{I}$ можно взять формулу

$$\forall w\ [F(w, y)\ \Rightarrow\ G(w, x)]. \tag{16}$$

Тогда формула $\mathbf{I}(n, e)$ как раз и будет истинной в точности для тех пар чисел $n$ и $e$, для которых выполнено неравенство $K_f(e) > n$, что и требовалось.

## Дорога ШЕНЯ (сообщена в 2007 г.)

Эта дорога основана на понятии программы вычислимой функции и на связанной с этим понятием одной из наиболее замечательных теорем теории алгоритмов — теореме о неподвижной точке. В этой теореме рассматриваются алгоритмы, преобразующие программы вычислимых функций в программы вычислимых функций; предполагается, что алгоритм приводит к результату в применении к любой программе. Теорема о неподвижной точке гласит, что каков бы ни был такой алгоритм, одна из программ непременно перейдёт в эквивалентную ей программу, то есть в программу той же самой функции. Теорема Гёделя получается из теоремы о неподвижной точке следующим образом. Строим такой алгоритм преобразования программ: для каждой программы $P$ алгоритм выдаёт в качестве результата $\mathbf{a}(P)$ первую из таких программ, неэквивалентность которых программе $P$ доказуема в рассматриваемом формальном языке. Если бы этот алгоритм давал результат на всех программах, то в силу теоремы о неподвижной точки нашлась бы программа $P_0$ с противоречивыми свойствами: $P_0$ и $\mathbf{a}(P_0)$ были бы эквивалентны и в то же самое время их неэквивалентность была бы доказуема. Поэтому наш алгоритм не для всех программ приводит к результату. А, значит, найдётся такая замечательная программа $Q$, что ни для какой программы нельзя доказать её неэквивалентность программе $Q$. Достаточно теперь взять любую программу, не эквивалентную программе $Q$, чтобы получить истинное, но не доказуемое утверждение.

Строгое изложение указанной идеи состоит из двух частей. В первой части интуитивное понятие программы заменяется некоторым математическим понятием, а теорема о неподвижной точке приобретает более точную формулировку. Во второй части уточняется переход от теоремы о неподвижной точке к Теореме Гёделя.

### Первая часть

В применении к выражениям, которые могут быть и не определены, вместо знака равенства $=$ используется *знак условного равенства* $\simeq$. Запись $A \simeq B$ означает, что выражения $A$ и $B$ одновременно определены и не определены и если определены, то их значения совпадают. Таким образом, утверждение $\forall x\ [x - x \simeq y - y]$ верно, а утверждение $\forall x\ [\frac{x}{x} \simeq \frac{y}{y}]$ неверно. Если не требовать, чтобы функции $\varphi$ и $\psi$ были определены для всех значений аргумента, их совпадение надо выразить так: $\forall x\ [\varphi(x) \simeq \psi(x)]$.

Говоря о программах вычислимых функций, мы имеем здесь в виду вычислимые функции одной переменной с натуральными аргументами и значениями. Чтобы перейти от интуитивного понятия программы к его точно описанному аналогу, укажем, какие свойства интуитивного понятия нам требуются. Прежде всего, мы предполагаем, что все программы записаны в виде слов в некотором алфавите (*алфавите программ*) и что в соответствующем словарном пространстве они образуют разрешимое множество. Раз множество программ разрешимо, то оно перечислимо и, будучи бесконечным, может быть вычислимо и без повторений занумеровано натуральными числами, так что существуют как алгоритм перехода от номера к программе с этим номером, так и алгоритм перехода от программы к её номеру. Обозначим через $p_n$ программу с номером $n$. Переходя от программ к их номерам, можно так переформулировать теорему о неподвижной точке: для всякой всюду определённой вычислимой функции $\chi$ из $\mathbf{N}$ в $\mathbf{N}$ найдётся такое число $q$, что $p_q$ и $p_{\chi(q)}$ эквивалентны, то есть суть программы одной и той же функции.

Мы предполагаем, что существует алгоритм, который для каждого $n$ и для каждого $x$ даёт результат применения программы $p_n$ ко входу $x$ и ничего не даёт, если результата применения $p_n$ к $x$ не существует. Этот алгоритм вычисляет вычислимую функцию $F$ от двух аргументов: $n$ и $x$. Таким образом, если закрепить в этой функции $F$ некоторое фиксированное число $n$ в качестве первого аргумента, то возникшая функция от второго (переменного) аргумента будет не что иное, как функция, вычисляемая программой $p_n$. Поэтому для функции $F$ выполняется такое

**свойство универсальности**: для всякой вычислимой функции $f$ из $\mathbf{N}$ в $\mathbf{N}$ существует такое число $n$, что для всякого $x$ имеет место условное равенство

$$f(x) \simeq F(n,x).$$

(В качестве $n$ достаточно взять номер программы, вычисляющей $f$.)

Для функции $F$ выполняется также

**свойство главности**: для всякой вычислимой функции $G$ из $\mathbf{N}\times\mathbf{N}$ в $\mathbf{N}$ существует такая всюду определённая вычислимая функция $\pi$ из $\mathbf{N}$ в $\mathbf{N}$, что

$$\forall n\ \forall x\ [G(n,x) \simeq F(\pi(n),x)].$$

(В качестве $\pi(n)$ надо взять номер программы, которая осуществляет вычисление функции от $x$ по следующей схеме: "к паре $(n,x)$ примени алгоритм вычисления функции $G$".)

Заметим, что из второго свойства вытекает первое, но мы разделили их для наглядности. Всякая вычислимая функция, обладающая свойством универсальности, называется *универсальной*. Всякая вычислимая универсальная функция, обладающая к тому же свойством главности, называется *главной универсальной*. Таким образом, описанная выше функция $F$ — одна из главных универсальных функций. В определении главной универсальной функции и отражаются существенные свойства программ.

Пусть $\Phi$ — произвольная числовая функция двух переменных. Число $n$ называется **номером** функции $f$ одной переменной **относительно** функции $\Phi$, коль скоро

$$\forall x\ [f(x) \simeq \Phi(n,x)].$$

Теорема о неподвижной точке приобретает теперь такую точную формулировку:

Пусть $\Phi$ — главная универсальная функция. Тогда для всякой всюду определённой вычислимой функции $\chi$ из $\mathbf{N}$ в $\mathbf{N}$ найдётся такое число $q$, что $q$ и $\chi(q)$ суть номера одной и той же функции относительно $\Phi$.

## Вторая часть

Фиксируем какую-то функцию $\Phi$ двух переменных. Два числа назовём *эквивалентными относительно* $\Phi$, если они являются номерами одной и той же функции. В случае вычислимости функции $\Phi$ утверждение об эквивалентности чисел $m$ и $n$ можно выразить в языке, то есть можно указать такую формулу с двумя параметрами, которая тогда и только тогда обращается в истину при подстановке чисел $m$ и $n$ вместо этих параметров, когда числа $m$ и $n$ эквивалентны. Укажем такую формулу. Поскольку $\Phi$ вычислима, её график перечислим и выражается некоторой формулой $A$ с параметрами $x$, $y$ и $z$, так что для любой тройки чисел $i,\ k,\ l$

$$A(i,k,l) \text{ истинна } \quad \text{тогда и только тогда, когда} \quad \Phi(i,k)=l.$$

Эквивалентность чисел $m$ и $n$ означает, что

$$\forall x\ [\Phi(m,x) \simeq \Phi(n,x)],$$

то есть, что для любой пары чисел $k$ и $l$

$$\Phi(m,k)=l \quad \text{тогда и только тогда, когда} \quad \Phi(n,k)=l.$$

Последняя равносильность выражается такой формулой:

$$[A(m,k,l) \Rightarrow A(n,k,l)]\ \&\ [A(n,k,l) \Rightarrow A(m,k,l)].$$

А эквивалентность чисел $m$ и $n$ выражается формулой

$$\forall y\ \forall z\ \{[A(m,y,z) \Rightarrow A(n,y,z)]\ \&\ [A(n,y,z) \Rightarrow A(m,y,z)]\}.$$

Обозначим эту формулу через $B(m,n)$.

Итак, формула $B(m,n)$ выражает эквивалентность чисел $m$ и $n$ относительно $\Phi$. Соответственно, неэквивалентность чисел $m$ и $n$ выражается формулой $\neg B(m,n)$.

Выберем теперь в качестве $\Phi$ какую-нибудь главную универсальную функцию. Расположим все доказуемые формулы в перечислимую последовательность. Рассмотрим алгоритм, который для каждого натурального $m$ сперва находит в последовательности доказуемых формул первую формулу вида $\neg B(m,n)$ с этим $m$ и каким-то $n$, а затем выдаёт это $n$ в качестве результата. Указанный алгоритм преобразования $m$ в $n$ вычисляет некоторую функцию $\chi$. График этой функции перечислим и потому выражается некоторой формулой $G$.

Покажем, что Функция $\chi$ определена не для всех натуральных чисел. Убедимся в этом способом от противного. Предположим, что функция $\chi$ определена для всех натуральных чисел. Тогда к ней можно применить теорему о неподвижной точке. Значит, существует такое число $m_0$, что числа $m_0$ и $\chi(m_0)$ эквивалентны и, стало быть, формула $B(m_0,\chi(m_0))$ истинна. По построению функции $\chi$, формула $\neg B(m_0,\chi(m_0))$ доказуема в формальном языки, значит, тоже истинна. Но формула не может быть истинной вместе со своим отрицанием.

Итак, $\chi$ не определена для некоторого натурального $q$. Это значит, что среди формул вида $\neg B(q,n)$ с этим $q$ и произвольным $n$ нет доказуемых формул.

В силу универсальности функции $\Phi$ имеется бесконечно много чисел, не эквивалентных числу $q$. Для любого такого $n$ формула $\neg B(q, n)$ окажется истинной, но не доказуемой. Что и даёт теорему Гёделя.

Упражнение. Как в формальном языке выразить теорему о неподвижной точке? (Проблема состоит в том, чтобы как-то выразить оборот “для всякой всюду определённой вычислимой функции $\chi$ одной переменной”.)