Chapter 2
Peano Arithmetic, Incompleteness

2.1 The language of PA

We have seen that some profound claims about the natural numbers are Π^0_1. On the other hand, the vocabulary of number theory allows us to formulate sentences that are not Π^0_1. Hilbert would say that at least some of these sentences lack content. We are going to consider a theory replete with such sentences. The theory is Peano Arithmetic (PA), a formalized version of elementary number theory.[1] A formal theory needs a formal language. The language of PA has the following vocabulary.

1. Two CONNECTIVES: '$-$' ("not"), '$\rightarrow$' ("if …then").
2. A QUANTIFIER: '$\forall$' ("for all").
3. The IDENTITY symbol: '$=$'.
4. Three FUNCTION symbols: 'S' ("successor"), '$+$' ("plus"), '$\cdot$' ("times").
5. One PROPER NAME: '0' ("zero").
6. Infinitely many VARIABLES: 'w', 'x', 'y', 'z', 'w_1', 'x_1', 'y_1', 'z_1', …
7. Two PARENTHESES: '(', ')'.

In everyday English, certain expressions refer to individual things or can so refer in appropriate contexts. Examples include proper names ('Kurt Gödel'), pronouns ('him'), and descriptions ('the second son of Marianne Gödel'). The TERMS of PA are the expressions that play this role in our formal language. We define them recursively as follows.

8. '0' is a term.
9. Every variable is a term.
10. If α and β are terms, then so are $\ulcorner S\alpha \urcorner$, $\ulcorner(\alpha + \beta)\urcorner$, and $\ulcorner(\alpha \cdot \beta)\urcorner$.

In everyday English, declarative sentences declare that something is the case ('Kurt Gödel was smart') or *could*, if placed in appropriate contexts, declare that

[1] Although the 'P' in 'PA' commemorates the mathematician GIUSEPPE PEANO (1858–1932), another mathematician, RICHARD DEDEKIND (1831–1916), deserves much of the credit. For some of the history, see Wang [11] (http://www.jstor.org/stable/2964176).

S. Pollard, *A Mathematical Prelude to the Philosophy of Mathematics*,
DOI: 10.1007/978-3-319-05816-0_2,

something is the case ('He was smart'). The FORMULAS of PA are the expressions that play this role in our formal language. We define them recursively as follows.

11. If α and β are terms, then $\ulcorner\alpha = \beta\urcorner$ is a formula.
12. If ϕ and ψ are formulas, then so are $\ulcorner-\phi\urcorner$ and $\ulcorner(\phi \to \psi)\urcorner$.
13. If α is a variable and ϕ is a formula, then $\ulcorner\forall\alpha\ \phi\urcorner$ is a formula.

Symbols are said to OCCUR in terms and formulas. For example, the function symbol '+' occurs in the term '$((x + y) + z)$'. In fact, it occurs twice: it has two OCCURRENCES. The left-hand parenthesis '(' also occurs twice, as does the right-hand parenthesis ')'. The variables 'x', 'y', and 'z' each occur once.

The symbols '$\ulcorner$' and '$\urcorner$' are known as CORNER QUOTES.[2] If α is a term, then $\ulcorner S\alpha\urcorner$ is the term consisting of an occurrence of the function symbol 'S' followed by an occurrence of the term α. If α is '0', then $\ulcorner S\alpha\urcorner$ is '$S0$'. Other corner quotations are to be understood similarly. Note that ordinary single quotation marks would not do what we intend here. For example, '$S\alpha$' is not a term of PA even if α is. This is because 'α' is not a term of PA. 'α' is an expression we use to discuss the terms of PA. To take a less technical example: 'David Hilbert' is an expression we use to designate a certain mathematician. To say that 'David Hilbert' was a mathematician is to say that David Hilbert's name was a mathematician. As far as we know, no mathematician has ever been David Hilbert's name (Such an arrangement is unlikely to be practical).

Exercise 2.1 *Suppose α is a PA-term. Is $\ulcorner\alpha S\urcorner$ a PA-term? Is $\ulcorner S\alpha\urcorner$ a PA-term? Does α occur in $\ulcorner S\alpha\urcorner$? Does α occur in '$S\alpha$'? Does 'α' occur in $\ulcorner S\alpha\urcorner$? Does '$\alpha$' occur in '$S\alpha$'?*

The SUB- FORMULAS of a formula ψ are ψ itself and any formulas that occur in ψ. For example, the formula

$$\forall x \forall y\ (x + Sy) = ((x + y) + S0)$$

has three sub-formulas. They are

$$(x + Sy) = ((x + y) + S0),$$

$$\forall y\ (x + Sy) = ((x + y) + S0),$$

and the whole formula itself. Any occurrence of a variable α within a sub-formula of ψ of the form $\ulcorner\forall\alpha\ \phi\urcorner$ is BOUND in ψ. All other occurrences are FREE in ψ. There are three bound occurrences of 'x' in

$$\forall x \forall y\ (x + Sy) = ((x + y) + S0).$$

[2] See Quine [9, pp. 33–37]. I will soon get sloppy and stop using the corner quotes (Absolute rigor does get a bit tedious). I thought, however, that you should know about them.

There are two free occurrences of 'x' in

$$\forall y\ (x + Sy) = ((x + y) + S0).$$

A free variable is like a pronoun without an antecedent. The expressions 'He is short' and '$x = 0$', outside of any appropriate context, make no definite claims and, so, are neither true nor false. (Who is *he* and what is x?)

Exercise 2.2 *Identify the bound occurrences of 'x', 'y', and 'z' in the following formula.*

$$\forall y \forall z(-\forall x - (z = Sy \rightarrow z = x) \rightarrow (x + y) = S(x \cdot SS0))$$

In our formal language, we understand SENTENCES to be formulas with no free occurrences of variables. This is a departure from everyday usage where 'He is short' counts as a declarative sentence even in contexts where the pronoun 'He' fails to refer to anyone and, so, is behaving like a free variable. Each sentence of PA makes a definite claim as soon as we supply PA with an interpretation (See Sect. 2.3 below).

2.2 The Axioms of PA

We will now use our formal language to describe the natural numbers (0, 1, 2, 3, ...). Actually, we will only describe how the natural numbers are related to one another: we will only describe the *structure* formed by the natural numbers or (speaking more circumspectly) the structure they *would* form if they were to exist. We do this by accepting certain sentences, the AXIOMS, without proof. We show that a sentence is a THEOREM by showing that it follows from the axioms. PA has two axioms characterizing the successor operator. The first says that 0 is not the immediate successor of any natural number ("Given any natural number x, it is not the case that 0 is identical to the immediate successor of x"). The second says that natural numbers with the same immediate successor are the same ("Given any natural numbers x and y, if the immediate successor of x is identical to the immediate successor of y, then x is identical to y").

S1 $\forall x - 0 = Sx$
S2 $\forall x \forall y(Sx = Sy \rightarrow x = y)$

PA has two axioms giving the recursive definition of addition and another two offering the recursive definition of multiplication.

A1 $\forall y\ (y + 0) = y$
A2 $\forall x \forall y\ (y + Sx) = S(y + x)$

M1 $\forall y\ (y \cdot 0) = 0$
M2 $\forall x \forall y\ (y \cdot Sx) = ((y \cdot x) + y)$

Finally, PA has infinitely many induction axioms, one for each formula with free occurrences of 'x' and 'y'. That is, if $\phi(x, y)$ is a formula of PA with free occurrences of 'x' and 'y', but no free occurrences of any other variable, if $\phi(0, y)$ is the result of replacing each free occurrence of 'x' in $\phi(x, y)$ with an occurrence of '0', and if $\phi(Sx, y)$ is the result of replacing each free occurrence of 'x' in $\phi(x, y)$ with an occurrence of 'Sx', then

$$\ulcorner \forall y(\phi(0, y) \rightarrow (\forall x(\phi(x, y) \rightarrow \phi(Sx, y)) \rightarrow \forall x\ \phi(x, y)))\urcorner$$

is an axiom of PA.

An example may clarify the connection between our induction axioms and the proofs by induction we did in Chap. 1. We can let $\phi(x, y)$ be any PA-formula with free occurrences only of 'x' and 'y'. If we let $\phi(x, y)$ be '$(y + x) = x$', we obtain the axiom

$$\forall y((y + 0) = 0 \rightarrow (\forall x((y + x) = x \rightarrow (y + Sx) = Sx) \rightarrow \forall x\ (y + x) = x))$$

This says the formula

$$((y + 0) = 0 \rightarrow (\forall x((y + x) = x \rightarrow (y + Sx) = Sx) \rightarrow \forall x\ (y + x) = x))$$

will be true no matter what y is. So, in particular, we are free to replace each free occurrence of 'y' with an occurrence of '0':

$$((0 + 0) = 0 \rightarrow (\forall x((0 + x) = x \rightarrow (0 + Sx) = Sx) \rightarrow \forall x\ (0 + x) = x))$$

This sentence says that you can show

$$\forall x\ (0 + x) = x$$

by first showing

$$(0 + 0) = 0$$

(to get through the first '$\rightarrow$') and then showing

$$\forall x((0 + x) = x \rightarrow (0 + Sx) = Sx)$$

(to get through the last '$\rightarrow$'). The goal here is to show that each natural number x is identical to $0 + x$. The first step is to show that 0 has this property: that 0 is identical to $0 + 0$. The next step is to show that this property is hereditary: that each natural number passes it on to its successor. You might think of the left-hand part of the formula

$$((0 + x) = x \to (0 + Sx) = Sx)$$

as an inductive hypothesis, while the right-hand part is the result to be obtained from that hypothesis.

Exercise 2.3 *Using A1, A2, and the induction axiom we have been discussing, prove informally that* $\forall x\ (0+x) = x$. (*I say "informally" because I have not introduced a formal deductive system of the sort you may have studied in a logic course. You still need to make good inferences when you do this exercise, but you can just go ahead and make those inferences without citing any explicit inference rules*).

2.3 Incompleteness 1: Compactness

As I mentioned above, a theorem of PA is a sentence in the language of PA that follows from the axioms of PA. A PROOF in PA is a demonstration that a PA-sentence does follow from the PA-axioms. A FORMAL LOGIC for PA will include a definition of 'proof' precise enough to yield a mechanical procedure for telling whether something is a proof. The logic of PA, CLASSICAL FIRST- ORDER LOGIC, was first formalized by GOTTLOB FREGE (1848–1925).[3] There are a variety of formalizations equivalent to Frege's. Assume we have picked one and suppose Γ is a set of sentences.

Definition 2.1 ψ is DERIVABLE from Γ if and only if there is a formal proof whose conclusion is ψ and whose premises are members of Γ. If ψ is derivable from Γ, we write: $\Gamma \vdash \psi$. If ψ is derivable from axioms of PA, we write: $\mathsf{PA} \vdash \psi$.

We could have characterized the axioms and theorems of PA without offering any hints about what those axioms and theorems *say*. This leaves us free to supply various readings or interpretations. In an INTERPRETATION of PA, we (1) specify the range of our bound variables, (2) assign an object from that range to '0', and (3) assign operations defined on that range to each of 'S', '+', and '$\cdot$'. If we are not feeling too adventurous we might (1) let our bound variables range over the natural numbers (so that we read '$\forall x$' as "for all natural numbers x"), (2) let '0' be our name for zero, and (3) assign the operations of immediate succession, addition, and multiplication to 'S', '+', and '$\cdot$'. This, after all, is the *intended* interpretation of PA. You will perhaps agree that all the axioms of PA come out true when so interpreted. An interpretation that makes all the axioms of PA true is said to be a MODEL of PA.

Here is an alternative interpretation. Let our bound variables range over zero and all the negative integers (so that we read '$\forall x$' as "for all non-positive integers x"). Let '0' be our name for zero. Assign the operation "minus one" to 'S' and the operation of addition to '+'. Read '$x \cdot y$' as "x times y times negative one". You might take a

[3] See Frege [2]; English translation in van Heijenoort [10, pp. 5–82]. If you have had a logic course, you have probably worked with a close cousin of Frege's system.

minute to confirm that all the axioms of PA come out true when interpreted in this way (So this is a model of PA).

Exercise 2.4 *Let our bound variables range over all the even natural numbers (including zero). Complete this interpretation in a way that makes the first six axioms of* PA *true.*[4]

I already noted that a proof is a demonstration that a sentence (the conclusion) follows from some sentences (the premises). I did not, however, define "follows from". That was probably fine: you probably already understood this sort of "following" well enough to make sense of the preceding discussion. Insisting on a definition of *everything* is a sure-fire way to block intellectual progress. Sometimes, however, a mathematical definition of a notion that is already well understood is a way of drawing that notion into the mathematical realm: making it the object of productive mathematical inquiry. That is the goal of the following definition.

Definition 2.2 ψ FOLLOWS FROM Γ if and only if it is logically impossible for an interpretation to make all the members of Γ true and ψ false. If ψ follows from Γ, we write: $\Gamma \vDash \psi$. If ψ follows from axioms of PA, we write: $\mathsf{PA} \vDash \psi$.

OK, we still have work to do: we do not yet have a mathematical definition of logical impossibility. Supplying such a definition is one of the most important jobs of set theory (You might think of set theory as an inventory of logically possible structures). We will leave a serious exposition of set theory for later chapters. For now, we will try to make do with a less systematic understanding of logical impossibility. Even without a full-blown theory, we have a good idea of how to establish that a situation is logically impossible: assume that the situation is real and derive a contradiction from that assumption.

There is another bit of work we will leave undone: we will not offer a mathematically precise account of what our logical symbols mean. This *can* be done (You may have already seen it done in a logic course. You saw a bit of it done in Chap. 1). We will just not be doing it here. If, in what follows, you need to figure out whether it is logically possible for an interpretation to make certain sentences true or false, you will have to draw on your rough-and-ready understanding of those sentences. I hope it is clear, for example, that no interpretation can make '$0 = 0$' false since that would require that the object assigned to '0' be distinct from the object assigned to '0'.

It turned out to be mathematically fruitful to give a mathematical definition of the "follows from" relation. In his doctoral dissertation of 1929, for example, Kurt Gödel proved that Frege's formalization of first-order logic is COMPLETE: any conclusion that follows from a set of premises is derivable from that set.[5]

Theorem 2.1 $\Gamma \vDash \psi$ *only if* $\Gamma \vdash \psi$.

---

[4] More modestly, we want an interpretation that *would* make the first six axioms true if there *were* such things as the natural numbers. This sort of conditional claim is what I will generally intend when I talk about an interpretation making certain sentences true.

[5] Gödel's dissertation is reprinted, with an English translation, in Gödel [7, pp. 60–101].

Frege's formalization of first-order logic is also SOUND: any conclusion that is derivable from a set of premises follows from that set.

Theorem 2.2 $\Gamma \vdash \psi$ *only if* $\Gamma \vDash \psi$.

Exercise 2.5 *Formal proofs are all finite: they consist of finitely many lines featuring only finitely many premises. Use this fact, and the preceding two theorems, to prove Gödel's* COMPACTNESS THEOREM*: if a formula follows from* Γ*, it follows from a finite subset of* Γ.

Exercise 2.6 *Suppose it is logically possible for* PA *to have a model. Show that* PA $\nvdash 0 \neq 0$.

Exercise 2.7 *Suppose* PA $\nvdash 0 \neq 0$. *Show that it is logically possible for* PA *to have a model.*

We are going to see that PA suffers from a particular form of incompleteness. First, though, we need to reflect on the relationship between numbers and numerals. The NUMERALS of PA are '0' and any terms consisting of an occurrence of '0' preceded by finitely many occurrences of 'S'. That is:

$$\text{`}0\text{'},\ \text{`}S0\text{'},\ \text{`}SS0\text{'},\ \text{`}SSS0\text{'},\ \ldots$$

We understand the natural numbers to be 0 and everything obtainable from 0 by finitely many applications of the successor operation S. So it follows from our conception of the natural numbers that each of them is named by a numeral of PA when these numerals are interpreted in the standard way (with '0' naming 0 and 'S' expressing the successor operation). For example, the number obtained from 0 by 15 application of S is named by the numeral consisting of an occurrence of '0' preceded by 15 occurrences of 'S'. The STANDARD numbers are 0 and all the numbers obtainable from 0 by finitely many applications of S: that is, exactly the numbers named by our numerals. It is part of our concept of the natural numbers that each of them is standard. So, if PA completely captures our concept of the natural numbers, the axioms of PA will rule out non-standard numbers: that is, it will be logically impossible for an interpretation to make all the axioms of PA true while including in the range of PA's bound variables an object that is not named by any numeral of PA.

At this point it is useful to introduce two new symbols: the proper name 'c' and the inequality symbol '$\neq$'. We use '$\neq$' to make our claims about non-identity a bit more readable:

$$\alpha \neq \beta \iff -\alpha = \beta.$$

You might think of 'c' as the name of a natural number (though I have not said which one). Now consider the following sequence of sentences.

$$c \neq 0, \quad c \neq S0, \quad c \neq SS0, \quad c \neq SSS0, \quad \ldots$$

Let C be the set of all these sentences. Each member of C says something compatible with our conception of the natural numbers. For example,

$$c \neq SSSSS0$$

makes the innocent claim that a certain (so far unidentified) natural number is distinct from five. Given any model of PA, we could make the above sentence true by assigning to 'c' an object in the range of our bound variables. Just let c be anything other than the object named by '$SSSSS0$'.

Exercise 2.8 *Suppose C^* is a proper subset of C. That is, every member of C^* is a member of C, but not every member of C is a member of C^*. Show the following: given any model of* PA*, we can make all the members of C^* true by assigning to 'c' an object in the range of our bound variables. You may assume the following: If α and β are* PA*-numerals and if a model of* PA *makes the equation $\ulcorner \alpha = \beta \urcorner$ true, then $\alpha = \beta$.* (*Can you see why this is so?*)

Let $\mathsf{PA} \cup C$ be the set consisting of the axioms of PA and the members of C. An interpretation of $\mathsf{PA} \cup C$ is an interpretation of PA that, in addition, assigns to 'c' an object in the range of PA's bound variables. We now adopt a premise for a conditional proof: that is, an argument intended to establish an "if ... then" statement. We need not believe that this premise is true. Our project is to see what follows from it.

Premise (for conditional proof): PA completely captures our concept of the natural numbers and, so, rules out non-standard numbers.

This would mean that a model of $\mathsf{PA} \cup C$ in which c behaves like a non-standard number is logically impossible. On the other hand, if we make all the members of C true, c will *have* to be non-standard, since c will be distinct from every number named by a numeral of PA. So our premise implies that it is logically impossible for an interpretation to make all the members of $\mathsf{PA} \cup C$ true. This means that

$$\mathsf{PA} \cup C \vDash \psi$$

no matter what ψ is. (Check the definition of 'follows from' to confirm this.) For example,

$$\mathsf{PA} \cup C \vDash 0 \neq 0.$$

So, by the Compactness Theorem, '$0 \neq 0$' follows from *finitely many* members of $\mathsf{PA} \cup C$. We may suppose, then, that

$$\mathsf{PA}^* \cup C^* \vDash 0 \neq 0$$

where PA^* is a finite set of PA-axioms and C^* is a finite subset of C. Any interpretation that makes all the members of PA true will make all the members of PA^* true and can easily be extended to make the finitely many members of C^* true (as you showed in the last exercise). Such an interpretation would have to make '$0 \neq 0$' true since '$0 \neq 0$' follows from $\mathsf{PA}^* \cup C^*$. But that is logically impossible. (It is logically impossible for '0' to name an object that is not the same object as itself.) So there can be no such interpretation. That is, PA is UNSATISFIABLE: it is logically impossible for an interpretation to make all the axioms of PA true; it is logically impossible for PA to have a model. So

$$\mathsf{PA} \vDash 0 \neq 0$$

and, hence, by the Completeness Theorem,

$$\mathsf{PA} \vdash 0 \neq 0.$$

That is, PA is INCONSISTENT.[6] We reached this conclusion by first supposing that PA completely captures our concept of the natural numbers. So our grand conclusion is: PA completely captures our concept of the natural numbers only if PA is inconsistent.

If PA is consistent, it is INCOMPLETE. (It is a bit ironic that the incompleteness of PA follows from Gödel's *completeness* theorem.) What is at issue here is *expressive* incompleteness. There are mathematically important properties of the natural numbers that cannot even be expressed in the language of PA. Even if we use infinitely many sentences of PA, we cannot assert that every natural number is standard nor, to take another example, can we assert that each natural number is greater than only finitely many natural numbers.

Exercise 2.9 *Sketch a compactness argument showing that* PA *allows for infinitely large numbers. It might be helpful to start by using the vocabulary of* PA *to define the "less than" symbol '$<$'. Feel free to define any other vocabulary you find useful. Feel free, too, to assume that* PA *is consistent. You may, if you wish, fill in all the details of the argument, but I am only asking for the general idea.*

2.4 Incompleteness 2: Representability

In the intended interpretation of PA, the PA-NUMERAL for a natural number n consists of an occurrence of '0' preceded by n occurrences of 'S'. When we use the informal variable 'n' to make general claims about natural numbers, we can use the informal variable '$\mathbf{n}$' to make general claims about the corresponding PA-numerals. For example, we might say: if n is a natural number, then the PA-numeral $\mathbf{n}$ consists of an occurrence of '0' preceded by n occurrences of 'S'.

---

[6] Here are three equivalent definitions of inconsistency: an inconsistent theory is one that proves a logical absurdity such as '$0 \neq 0$'; an inconsistent theory is one that proves a sentence ϕ and its negation $-\phi$; an inconsistent theory is one that proves every sentence in its language.

Suppose f is a one-place, primitive recursive, characteristic function. By "primitive recursive", we mean essentially what we meant in the preceding chapter with any modifications necessary to accommodate the number zero. A one-place function takes natural numbers one at a time as inputs. A characteristic function has only two possible outputs: zero and one.

Suppose $\phi(x)$ is a PA-formula with free occurrences of 'x' and no free occurrences of any other variable. Then $\phi(x)$ is said to REPRESENT f in PA if it has the following properties. If n is a natural number and $f(n) = 1$, then PA proves the sentence that results when we replace every free occurrence of 'x' in $\phi(x)$ with an occurrence of the PA-numeral for n. That is:

$$f(n) = 1 \implies \mathsf{PA} \vdash \phi(\mathbf{n}).$$

If n is a natural number and $f(n) = 0$, then PA proves the sentence that results when we replace every free occurrence of 'x' in $-\phi(x)$ with an occurrence of the PA-numeral for n. That is:

$$f(n) = 0 \implies \mathsf{PA} \vdash -\phi(\mathbf{n}).$$

If f is represented in PA by a PA-formula, then we say that f is REPRESENTABLE in PA.

Each characteristic function answers a yes-or-no question about the natural numbers. It answers "yes" by returning the value 1. It answers "no" by returning the value 0. (In Chap. 1, || was "yes", while | was "no".) Suppose the characteristic function f is represented in PA by the formula $\phi(x)$. By feeding f a number n, we can learn whether PA thinks n has the property expressed by $\phi(x)$. f will answer "yes" only if PA proves $\phi(\mathbf{n})$. f will answer "no" only if PA refutes $\phi(\mathbf{n})$.

Here is an example. We can infer from Definition 1.14, that there is a primitive recursive characteristic function $\mathfrak{O}$ that behaves as follows.

$$\mathfrak{O}(n) = \begin{cases} 1 & \text{if } n \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

$\mathfrak{O}$ tells us whether the numbers we feed it are odd. If we want to find a PA-formula that represents f, we should look for one that somehow expresses the property of oddness. If $\mathfrak{O}$ were represented in PA by $\phi(x)$, then PA would prove each of the sentences

$$\phi(S0),\ \phi(SSS0),\ \phi(SSSSS0),\ \phi(SSSSSSS0), \ldots$$

and would disprove each of the sentences

$$\phi(0),\ \phi(SS0),\ \phi(SSSS0),\ \phi(SSSSSS0), \ldots$$

That is, PA would prove $\phi(\mathbf{n})$ whenever n is odd and would prove $-\phi(\mathbf{n})$ whenever n is even. Can we identify such a formula $\phi(x)$? Well, does the vocabulary of PA allow us to say that a number x is odd? Of course it does. We just say that x is not a multiple of two:

$$\forall y\ (y \cdot SS0) \neq x.$$

As it turn out, this formula really does represent f because PA proves each of the sentences

$$\forall y\ (y \cdot SS0) \neq S0,\ \forall y\ (y \cdot SS0) \neq SSS0,\ \forall y\ (y \cdot SS0) \neq SSSSS0,\ \ldots$$

and disproves each of the sentences

$$\forall y\ (y \cdot SS0) \neq 0,\ \forall y\ (y \cdot SS0) \neq SS0,\ \forall y\ (y \cdot SS0) \neq SSSS0,\ \ldots$$

So, when we offer f the number n and f responds "yes", we do not just learn that n is odd: we learn that PA thinks n is odd. When we offer f the number n and f responds "no", we do not just learn that n is even: we learn that PA thinks n is even.

Here is another example. We know from Definition 1.15, that there is a primitive recursive characteristic function $\mathfrak{P}$ that behaves as follows.

$$\mathfrak{P}(n) = \begin{cases} 1 & \text{if } n \text{ is prime} \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 2.10 *Identify a* PA*-formula that represents* $\mathfrak{P}$ *in* PA.

It is no fluke that $\mathfrak{O}$ and $\mathfrak{P}$ are representable in PA. Gödel proved in 1930 that *all* of our primitive recursive functions f are representable in PA.[7] So if

$$f(\mathfrak{a}) = 1$$

is a true Π^0_1 sentence, there is a formula $\phi(x)$ that represents f in PA and, so, PA proves each of the sentences

$$\phi(0),\ \phi(S0),\ \phi(SS0),\ \phi(SSS0),\ \ldots$$

Since PA does not guarantee that

$$0,\ S0,\ SS0,\ SSS0,\ \ldots$$

are the only objects in the range of its bound variables, we should hesitate to infer that PA will prove

---

[7] See Gödel [6]; English translation in Gödel [7, pp. 145–195], and van Heijenoort [10, pp. 596–616].

$$\forall x\ \phi(x).$$

Indeed, Gödel identified primitive recursive functions f represented by PA-formulas $\phi(x)$ such that, if PA is consistent, it is true that $f(\mathfrak{a}) = 1$, but $\forall x\ \phi(x)$ is not provable in PA. That is, Gödel actually identified formulas $\phi(x)$ with the following odd property. PA proves each of the sentences

$$\phi(0),\ \phi(S0),\ \phi(SS0),\ \phi(SSS0),\ \ldots$$

and, so, confirms that each *standard* number satisfies $\phi(x)$. Yet, if PA is consistent, PA is unable to confirm that every number, every object in the range of its bound variables, satisfies $\phi(x)$ (If PA is *inconsistent* it "confirms" everything: every PA-sentence is a PA-theorem).

Suppose, on the other hand, that $f(\mathfrak{a}) = 1$ and PA $\vdash -\forall x\ \phi(x)$. Then, by Theorem 2.2, PA $\models -\forall x\ \phi(x)$ and, hence, every model of PA makes $-\forall x\ \phi(x)$ true (since no interpretation makes all the PA-axioms true and $-\forall x\ \phi(x)$ false). So every model of PA makes $\forall x\ \phi(x)$ false. However, since $\phi(x)$ represents f in PA, Theorem 2.2 implies that every model of PA makes each of the sentences

$$\phi(0),\ \phi(S0),\ \phi(SS0),\ \phi(SSS0),\ \ldots$$

true. This would mean that every model of PA includes a non-standard number in the range of PA's bound variables (a number that makes $\forall x\ \phi(x)$ false). That is, PA would have no STANDARD MODELS and, hence, would be logically incompatible with our concept of the natural numbers. Our grand conclusion: if PA is compatible with our concept of the natural numbers, then PA does not *disprove* $\forall x\ \phi(x)$.

Suppose, now, it is logically possible for PA to have a standard model (a model in which every object in the range of PA's bound variables is named by a PA-numeral). Then Gödel has shown us how to identify primitive recursive functions f represented by PA-formulas $\phi(x)$ such that (1) it is true that $f(\mathfrak{a}) = 1$, but (2) $\forall x\ \phi(x)$ is neither provable nor refutable in PA. PA-sentences that are neither provable nor refutable in PA are said to be UNDECIDABLE in PA. We shall now consider what sort of an f would allow us to establish undecidability.

We first note that we can use natural numbers to *code* formulas of PA. Let's code some of the symbols of PA as follows.

$-$	$\rightarrow$	$\forall$	$=$	x	y	S	0	$+$	$\cdot$	(	)
1	2	3	4	5	6	7	8	9	10	11	12

Now consider the PA-sentence '$-0 = 0$'. To code this sentence, first replace each symbol with its code:

$$\begin{array}{cccc} - & 0 & = & 0 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 8 & 4 & 8 \end{array}$$

Now take the product of the first four prime numbers raised to the powers 1, 8, 4, 8:

$$2^1 \times 3^8 \times 5^4 \times 7^8.$$

This number

$$47{,}278{,}574{,}201{,}250$$

is the GÖDEL NUMBER of the PA-sentence '$-0 = 0$'.[8] Working the other direction, if you were presented with the number 47,278,574,201,250, you could ask your calculator or computer to factor it. Then, noting the exponents and consulting our table of symbol codes, you could reconstruct the PA-sentence '$-0 = 0$'.

Exercise 2.11 *Decode the Gödel number* 2,349,101,964,825,000.

Exercise 2.12 *Decode the Gödel number*

$$96{,}860{,}719{,}328{,}790{,}117{,}762{,}174{,}283{,}536{,}741{,}369{,}039{,}814{,}728{,}960{,}000.$$

We can use Gödel numbering to code sequences of PA-formulas. For example, if ϕ and ψ are PA-formulas with codes $\#\phi$ and $\#\psi$, we can let

$$2^{\#\phi} \times 3^{\#\psi}$$

code the ordered pair $< \phi, \psi >$. (We already used this trick in Chap. 1.) We can use similar techniques to code formalized PA-proofs. Suppose we have done so. Gödel figured out how to identify a primitive recursive function $\mathfrak{g}$, represented in PA by a PA-formula $\gamma(x)$, that behaves as follows.[9]

---

[8] The corresponding PA-numeral consists of a single occurrence of '0' preceded by 47,278,574,201,250 occurrences of 'S'. If you were to produce a token of this numeral, it would be about 100 million km long. That is about two thirds the distance from the Earth to the Sun or about 2,500 times the circumference of the Earth. This suggests that it may be naive to think of PA-numerals as actual physical objects.

[9] For a readable discussion of Gödel's construction, see Nagel and Newman [8]. Another helpful resource on this and other issues of interest to us is George and Velleman [5]. It might help you wrap your brain around Gödel's proof if you read $\gamma(\mathbf{n})$ as "n does not code a PA-proof of **G**" where **G** is a certain extra-special sentence of PA. Then $\forall x\ \gamma(x)$, the universal generalization of $\gamma(\mathbf{n})$, says that no natural number codes a PA-proof of **G**. Now it so happens that $\forall x\ \gamma(x)$ *is* **G**. So **G** says of itself that it is not provable in PA. A PA-proof of **G** would prove that **G** is not provable in PA: a strange situation, to say the least. Of course, all this is a bit sloppy. $\gamma(\mathbf{n})$ does not say *anything* unless we interpret it. Furthermore, under the intended interpretation it does not say anything about PA-proofs: it only refers to natural numbers. But the road to clarity is sometimes paved with slop.

$$\mathfrak{g}(n) = \begin{cases} 1 & \text{if } n \text{ does not code a proof in } \mathsf{PA} \text{ of } \forall x\ \gamma(x) \\ 0 & \text{otherwise.} \end{cases}$$

Since $\gamma(x)$ represents $\mathfrak{g}$ in PA we also know:

$$\mathfrak{g}(n) = 1 \implies \mathsf{PA} \vdash \gamma(\mathbf{n})$$

$$\mathfrak{g}(n) = 0 \implies \mathsf{PA} \vdash -\gamma(\mathbf{n}).$$

If we offer $\mathfrak{g}$ the number n and $\mathfrak{g}$ responds "yes", then n does *not* code a PA-proof of the generalization $\forall x\ \gamma(x)$, but PA does prove the instance $\gamma(\mathbf{n})$. The other case is more interesting: if we offer $\mathfrak{g}$ the number n and $\mathfrak{g}$ responds "no", then n codes a PA-proof of $\forall x\ \gamma(x)$ and PA proves $-\gamma(\mathbf{n})$. But it is not coherent to say that every number has a certain property and, also, that a particular number does not. So if $\mathfrak{g}$ were ever to say "no", that would mean PA is inconsistent. Let us run through this argument more carefully. Suppose $\mathsf{PA} \vdash \forall x\ \gamma(x)$. Then we can pick a natural number k that codes a PA-proof of $\forall x\ \gamma(x)$. Note that $\mathfrak{g}(k) = 0$. So, since $\gamma(x)$ represents $\mathfrak{g}$, $\mathsf{PA} \vdash -\gamma(\mathbf{k})$. But, since $\forall x\ \gamma(x)$ is derivable in PA, $\mathsf{PA} \vdash \gamma(\mathbf{k})$. That is, PA is inconsistent. Our conclusion: if PA is consistent, then PA does not prove $\forall x\ \gamma(x)$. Suppose, now, it is logically possible for PA to have a standard model. Then, by Exercise 2.6, PA is consistent. So no natural number codes a PA-proof of $\forall x\ \gamma(x)$ and, hence, the Π^0_1-sentence

$$\mathfrak{g}(\mathfrak{a}) = 1$$

is true. So PA proves each of the sentences

$$\gamma(0),\ \gamma(S0),\ \gamma(SS0),\ \gamma(SSS0),\ \ldots$$

and, hence, by our earlier reasoning, PA does not disprove $\forall x\ \gamma(x)$ (since, otherwise, every model of PA would have to feature a non-standard number witnessing to the falsehood of $\forall x\ \gamma(x)$). Our grand conclusion: if it logically possible for PA to have a standard model, then $\forall x\ \gamma(x)$ is undecidable (neither provable nor refutable) in PA.

We earlier saw that PA suffers from a kind of *expressive* incompleteness: there are mathematically important properties of the natural numbers that cannot be expressed in the language of PA. We now see that PA suffers from a kind of *proof-theoretic* incompleteness: there are questions expressible in the language of PA that cannot be settled by a proof or refutation in PA. That is, there are cases where PA's expressive resources are up to snuff, but its capacity to supply proofs is not.

Here is another example of an undecidable sentence. There is a primitive recursive function $\mathfrak{s}$, represented in PA by a PA-formula $\sigma(x)$, that behaves as follows.

$$\mathfrak{s}(n) = \begin{cases} 1 & \text{if } n \text{ does not code a proof in } \mathsf{PA} \text{ that } 0 \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose PA is consistent. That is, suppose PA does not prove the absurdity '$0 \neq 0$'. Then no natural number codes a PA-proof of '$0 \neq 0$' and, hence, the Π^0_1-sentence

$$\mathfrak{s}(\mathfrak{a}) = 1$$

is true. So PA proves each of the sentences

$$\sigma(0),\ \sigma(S0),\ \sigma(SS0),\ \sigma(SSS0),\ \ldots$$

and, hence, if it is logically possible for PA to have a standard model, PA does not disprove $\forall x\ \sigma(x)$. It turns out that PA does prove

$$(\forall x\ \sigma(x) \rightarrow \forall x\ \gamma(x)).$$

So, if PA proved $\forall x\ \sigma(x)$, it would also prove $\forall x\ \gamma(x)$. We conclude: if it is logically possible for PA to have a standard model, then $\forall x\ \sigma(x)$ is undecidable in PA.

Suppose there is a PA-formula $\chi(x)$ that somehow expresses the idea that x codes a PA-proof of '$0 \neq 0$'. If $\chi(x)$ is really doing a good job of expressing that idea inside PA, we would expect that, for each natural number n, if n does not code a PA-proof of '$0 \neq 0$', then PA $\vdash -\chi(\mathbf{n})$ (that is, PA will refute the sentence asserting that n *does* do the coding). Suppose this is the case.

Exercise 2.13 *Prove that:* PA $\vdash \forall x\ -\chi(x)$ *only if* PA *is inconsistent.* (*You might start by considering the relationship between the formula* $-\chi(x)$ *and the function* $\mathfrak{s}$.)

Exercise 2.14 *Suppose that, for each natural number* n, PA $\vdash -\chi(\mathbf{n})$ *only if* n *does not code a* PA*-proof of* '$0 \neq 0$'. *Show that this implies the consistency of* PA.

Exercise 2.13 shows: if PA allowed us to *say* that no natural number codes a PA-proof of '$0 \neq 0$', PA would not allow us to *prove* this unless it allowed us to prove *everything*. So we will not be able to use a PA-proof to demonstrate the consistency of PA. More generally, we will not be able to prove the consistency of PA using methods formalizable in PA.

Note that a PA-proof of PA's consistency would not be as pointless as it might first appear. The PA-proof would use only finitely many axioms of PA: we would be relying on only finitely many axioms to show that no combination of the infinitely many PA-axioms proves an absurdity. Those finitely many axioms might have had some property that made their consistency evident or, at least, more evident than the consistency of PA as a whole. So the PA-proof might have provided a non-circular reason for believing PA consistent. Alas, we now recognize that this is not to be.

2.5 Why Fret About Consistency?

Hilbert thought that mathematicians frequently devote time and talent to proofs of sentences that have no content: sentences that may seem to state something, but really state nothing. He thought it important to *justify* this practice in some way. He recognized that *consistency proofs* could provide an especially powerful justification. Here is why.

Suppose f is a primitive recursive characteristic function represented in PA by $\phi(x)$. If $f(n) = 0$, then $\mathsf{PA} \vdash -\phi(\mathbf{n})$. So if PA is consistent and $\mathsf{PA} \vdash \phi(\mathbf{n})$, then $f(n) = 1$ (since, otherwise, PA would prove both $\phi(\mathbf{n})$ and $-\phi(\mathbf{n})$). So if PA is consistent and $\mathsf{PA} \vdash \forall x\ \phi(x)$, then

$$f(0) = 1, \quad f(1) = 1, \quad f(2) = 1, \quad \ldots$$

are all true and, indeed, the Π_1^0 sentence

$$f(\mathfrak{a}) = 1$$

is true. If we are convinced that PA is consistent, we can feel free to use the machinery of PA to verify Π_1^0 sentences. We could think of PA as a trustworthy oracle. If the great oracle PA says that $\forall x\ \phi(x)$ is true, then that settles it: it really is true that $f(\mathfrak{a}) = 1$. The same argument applies to any theory in which all the PA-axioms are derivable: if the theory is consistent, we can use it to verify Π_1^0 statements. Let your imagination run wild: as long as your fanciful tales are not fundamentally incoherent, we can use them to verify propositions such as Fermat's Last Theorem and Goldbach's Conjecture. What matters is consistency, not truth.

If we have good reason to believe PA consistent, then we have good reason to use PA to prove theorems. If a theorem itself corresponds to a Π_1^0 sentence, we will verify that sentence. Otherwise, we will obtain results that (meaningless or not) may contribute to other proofs and, so, help us verify Π_1^0 sentences. Note that we need a *good reason* for believing PA consistent. A *mathematical proof* would certainly be a good reason. I note, though, that not every good reason in mathematics is supplied by a proof. (We might have very good reasons for adopting axioms, though axioms are statements we accept without proof.) Gerhard Gentzen (1909–1945) *did* prove the consistency of PA using techniques not formalizable in PA.[10] If you are already convinced that each axiom of PA is a true statement about the natural numbers, you may feel no need for such a proof. After all, a collection of *true* sentences cannot be inconsistent. Indeed, sentences that *could* all be true cannot be inconsistent. I suppose a reasonable person could, nonetheless, question the consistency of PA. However, the logician Solomon Feferman reports that, among today's mathematicians, "the number who doubt that PA is consistent is vanishingly small".[11]

[10] See Gentzen [3]; English translation in Gentzen [4, pp. 132–213].

[11] See Feferman [1, p. 192].

As I already mentioned, the above reasoning applies to any theory that extends PA, any theory in which all the PA-axioms are derivable. If we have good reason to believe such a theory consistent, we have good reason to use the theory to prove theorems. In the next chapter, we begin our study of an important family of such theories: set theories.

2.6 Solutions of Odd-Numbered Exercises

2.1 No. Yes. Yes. No. No. Yes.

2.3 A1 says that, no matter what y is, $(y + 0) = y$. So, in particular, $(0 + 0) = 0$. Suppose, as an inductive hypothesis, that $(0 + x) = x$. Then $S(0 + x) = Sx$. A2 says that, no matter what x and y are, $(y + Sx) = S(y + x)$. So, in particular, $(0 + Sx) = S(0 + x)$ and, hence, $(0 + Sx) = Sx$. We have shown

$$((0 + x) = x \rightarrow (0 + Sx) = Sx).$$

Since we have not used any special information about x, x could be *anything*. That is, $\forall x((0 + x) = x \rightarrow (0 + Sx) = Sx)$. One of our induction axioms assures us that

$$((0 + 0) = 0 \rightarrow (\forall x((0 + x) = x \rightarrow (0 + Sx) = Sx) \rightarrow \forall x\ (0 + x) = x)).$$

So $\forall x(0 + x) = x$.

2.5 Suppose $\Gamma \vDash \psi$. Then, by the completeness theorem, $\Gamma \vdash \psi$. That is, there is a formal proof whose conclusion is ψ and whose premises are members of Γ. Let the members of Γ' be the finitely many members of Γ that appear in the proof. Then $\Gamma' \vdash \psi$ and, hence, by the soundness theorem, $\Gamma' \vDash \psi$.

2.7 Suppose PA $\nvdash 0 \neq 0$. Then, by the completeness theorem, PA $\nvDash 0 \neq 0$. That is, it is logically possible for an interpretation to make all the axioms of PA true and '$0 \neq 0$' false.

2.9 We introduce the "less than" symbol by stipulating that $\ulcorner\alpha < \beta\urcorner$ is an abbreviation of $\ulcorner-\forall z\ (\alpha + Sz) \neq \beta\urcorner$. (The idea is that $x < y$ if and only if you can get to y by adding some non-zero number to x.) Let the members of C be the sentences

$$0 < c, S0 < c, SS0 < c, SSS0 < c, \ldots$$

Let C^* be a finite subset of C. Each member of C^* is an inequality $\ulcorner\alpha < c\urcorner$ where α is a PA-numeral. Since C^* is finite we can pick a PA-numeral β longer than all those α's. We are assuming that PA is consistent. So, by the completeness theorem and Exercise 2.7, it is logically possible for PA to have a model. Suppose we have picked such a model. That model will assign an object to β. Extend the model by assigning the same object to 'c'. Our new model will make $\ulcorner c = \beta \urcorner$ true. If our new model makes an inequality $\ulcorner \alpha < c \urcorner$ in C^* false, then (since our model thinks 'c' and β name the same thing) it will make the inequality $\ulcorner \alpha < \beta \urcorner$ false where α and β are PA-numerals, the latter longer than the former. We can show that this is impossible. (You might enjoy working out the details.) So, in fact, our new model makes all the members of C^* true. More generally, if C^* is any finite subset of C, it is logically possible for $\mathsf{PA} \cup C^*$ to have a model. If every member of C were true, then c would exceed every finite value and, so, would deserve to be called infinitely large. Suppose PA rules out infinitely large numbers. Then the combination $\mathsf{PA} \cup C$ is incoherent and, so, can have no models. Since no models are possible,

$$\mathsf{PA} \cup C \vDash 0 \neq 0.$$

Use the compactness theorem to pick a finite subset of C such that

$$\mathsf{PA} \cup C^* \vDash 0 \neq 0.$$

Then $\mathsf{PA} \cup C^*$ cannot have a model—contrary to our earlier result. So we were wrong to suppose that PA rules out infinitely large numbers.

2.11 $\forall x \; x = x$.

2.13 First note the following

$$\begin{aligned} \mathfrak{s}(n) = 1 &\Longrightarrow n \text{ does not code a proof in } \mathsf{PA} \text{ that } 0 \neq 0 \\ &\Longrightarrow \mathsf{PA} \vdash -\chi(\mathbf{n}) \\ \mathfrak{s}(n) = 0 &\Longrightarrow n \text{ codes a proof in } \mathsf{PA} \text{ that } 0 \neq 0 \\ &\Longrightarrow \mathsf{PA} \text{ proves every sentence of } \mathsf{PA} \\ &\Longrightarrow \mathsf{PA} \vdash - - \chi(\mathbf{n}). \end{aligned}$$

So $-\chi(x)$ represents $\mathfrak{s}$ in PA and, hence, by our earlier reasoning, if PA proved $\forall x \; - \chi(x)$, it would also prove $\forall x \; \gamma(x)$. But PA does *that* only if it is inconsistent.

References

1. Feferman, S. (1998). *In the light of logic*. New York: Oxford University Press.
2. Frege, G. (1879). *Begriffsschrift*. Halle: Louis Nebert.
3. Gentzen, G. (1936). Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, *112*, 493–565.
4. Gentzen, G. (1969). *The collected papers of Gerhard Gentzen*. Amsterdam: North-Holland.
5. George, A., & Velleman, D. J. (2002). *Philosophies of mathematics*. Oxford: Blackwell.

6. Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, *38*, 173–98.
7. Gödel, K. (1986). *Collected works* (Vol. 1). New York: Oxford University Press.
8. Nagel, E., & Newman, J. R. (1958). *Gödel's proof*. New York: New York University Press.
9. Quine, W. V. O. (1981). *Mathematical logic*. Cambridge, MA: Harvard University Press.
10. van Heijenoort, J. (Ed.). (1967). *From Frege to Gödel*. Cambridge, MA: Harvard University Press.
11. Wang, H. (1957). The axiomatization of arithmetic. *Journal of Symbolic Logic*, *22*, 145–58.

http://www.springer.com/978-3-319-05815-3

A Mathematical Prelude to the Philosophy of Mathematics
Pollard, S.
2014, XI, 202 p., Hardcover
ISBN: 978-3-319-05815-3