

GÖDEL'S THEOREM AND PARACONSISTENCY

Graham Priest¹

1. To Begin With

1.1 Introduction. Much has been written about Gödel's First Incompleteness Theorem. Nearly always, this is on the assumption that the logic of the theory of arithmetic in question is classical—or at least intuitionistic—logic. Rarely is the possibility that a paraconsistent logic is used explored. Yet the use of a paraconsistent logic puts a distinctively new spin on matters. The point of the present paper is to spell some of this out.

I shall not assume that people are familiar with the elements of paraconsistent logic—or at least, of the paraconsistent logic I will deploy. So before we get to paraconsistent arithmetic, I will spell this out. We will then be in a position to bring this to bear on Gödel's Theorem. Mostly, what I say will be of a technical nature; but I will permit myself a few philosophical comments towards the end.²

1.2 Gödel's Theorem. But first, some background.

If one looks at textbook statements of Gödel's Theorem, one will normally find something like this:

- Any axiomatic theory of arithmetic, with appropriate expressive capabilities, is incomplete—in the sense that there is something true (in the standard model of arithmetic) that cannot be proved.

This is inaccurate. What the proof of the theorem actually shows is this:

- Any axiomatic theory of arithmetic, with appropriate expressive capabilities, is *either* incomplete—in that sense—*or* inconsistent.

Of course, the inconsistent possibility is usually ignored, since in classical logic inconsistency implies triviality, and so this possibility is uninteresting. With a paraconsistent logic, it is, however.

It might be thought that the proof of the Theorem depends on classical logic. It does not. In fact, it makes virtually no assumptions about the underlying logic of the theory. Let me show this.

The theorem can be proved for any theory, T , which satisfies the following four conditions:

¹ Departments of Philosophy, the CUNY Graduate Center, and the University of Melbourne. E-mail: priest.graham@gmail.com.

² This is a written up version of a talk that has been given in various places, including a meeting of the Swiss Graduate Society for Logic and the Philosophy of Science (Bern University), and the conferences Ouroboros 2018: Formal Criteria of Self-Reference in Mathematics and Philosophy (University of Bonn), and As Formas da Razão (Universidade de Brasília).

- [A] Gödel codes are assigned to syntactic entities, such as formulas and proofs. If n is a number, write its numeral as \mathbf{n} . If A is a formula with code n , write $\langle A \rangle$ for \mathbf{n} .
- [B] There is a formula with two free variables, $B(x, y)$, which defines the proof relation of T . That is:
 - (i) if n is the code of a proof of A in T then $B(\mathbf{n}, \langle A \rangle)$ is true in the standard model
 - (ii) if n is not the code of a proof of A in T then $\neg B(\mathbf{n}, \langle A \rangle)$ is true in the standard model
- [C] Define $Prov\ y$ as $\exists x B(x, y)$. Then $Prov$ is a proof predicate for T . That is:
 - if $T \vdash A$ then $T \vdash Prov\langle A \rangle$
- [D] There is a formula, G , of the form $\neg Prov\langle G \rangle$.³

Given these assumptions, we can now establish Gödel's Theorem in the disjunctive form in which I stated it.

- By [D], if $T \vdash G$ then $T \vdash \neg Prov\langle G \rangle$.
- By [C], $T \vdash G$ then $T \vdash Prov\langle G \rangle$.
- Hence, if $T \vdash G$, T is inconsistent.

Suppose that T is consistent.

- Then $T \not\vdash G$.
- That is, no number is the code of a proof of G .
- By [Bii], for any n , $\neg B(\mathbf{n}, \langle G \rangle)$ is true in the standard model.
- Hence, $\forall x \neg B(x, \langle G \rangle)$ is true in the standard model.
 - As, then, is $\neg \exists x B(x, \langle G \rangle)$.
 - That is, $\neg Prov\langle G \rangle$.
 - That is, G .
- Hence, G is true in the standard model.

So if T is consistent, it is incomplete. Contrapositively, if T is complete, it is inconsistent.

Notice that virtually nothing is assumed in this argument about the logic of T .

2. Paraconsistent Logic

2.1 Semantics. Let us now turn to details concerning paraconsistent logic. The paraconsistent logic in question is LP . This may be set up in many ways. For our purposes, the following is most the appropriate.⁴

The language is a standard first-order language. We take $A \supset B$ to be defined in the usual way, as $\neg A \vee B$. An interpretation for the language is a structure $\mathfrak{A} = \langle D, \delta \rangle$ such that:

³ Standard proofs of the theorem usually do not deliver quite this. What they deliver is a formula, G , such that $G \equiv \neg Prov\langle G \rangle$ (and in a paraconsistent context, \equiv does not detach). However, what follows can be modified to take account of this fact with only minor changes. In any case, there are various ways to obtain a literal identity. My favourite is as follows. (I'm pretty sure that the idea is due to Saul Kripke, though I forget where I first learned it.) Extend the language of arithmetic by a new constant, \mathbf{k} . Assign Gödel codes to this language in the usual way. Let the code of $\neg \exists x Prov\ \mathbf{k}$ be k . Now, consider an interpretation of this language which is the same as the standard interpretation, except that the denotation of \mathbf{k} is k . In this interpretation, $\mathbf{k} = \langle \neg \exists x Prov\ \mathbf{k} \rangle$.

⁴ See Priest (2006a), ch. 5, and Priest (2008), chs. 7, 21.

- D is a non-empty domain (of quantification).
- For every constant, c , $\delta(c) \in D$.
- For every n -place function symbol, f_n , $\delta(f_n)$ is a function from D^n to D .
- For every n -place predicate, P_n , $\delta(P_n)$ is a pair, $\langle \delta^+(P_n), \delta^-(P_n) \rangle$ such that:
 - $\delta^+(P_n) \cup \delta^-(P_n) = D^n$
- $\delta^+(=) = \{ \langle d, d \rangle : d \in D \}$

Let $\mathfrak{A} \models^+ A$ and $\mathfrak{A} \models^- A$ mean that A is true and false in \mathfrak{A} , respectively. Then, given any interpretation, \mathfrak{A} :

- $\delta(f_n t_1 \dots t_n) = \delta(f_n)(\delta(t_1), \dots, \delta(t_n))$
- $\mathfrak{A} \models^+ P_n t_1 \dots t_n$ iff $\langle \delta(t_1), \dots, \delta(t_n) \rangle \in \delta^+(P_n)$
- $\mathfrak{A} \models^- P_n t_1 \dots t_n$ iff $\langle \delta(t_1), \dots, \delta(t_n) \rangle \in \delta^-(P_n)$
- $\mathfrak{A} \models^+ \neg A$ iff $\mathfrak{A} \models^- A$
- $\mathfrak{A} \models^- \neg A$ iff $\mathfrak{A} \models^+ A$
- $\mathfrak{A} \models^+ A \wedge B$ iff $\mathfrak{A} \models^+ A$ and $\mathfrak{A} \models^+ B$
- $\mathfrak{A} \models^- A \wedge B$ iff $\mathfrak{A} \models^- A$ or $\mathfrak{A} \models^- B$
- $\mathfrak{A} \models^+ A \vee B$ iff $\mathfrak{A} \models^+ A$ or $\mathfrak{A} \models^+ B$
- $\mathfrak{A} \models^- A \vee B$ iff $\mathfrak{A} \models^- A$ and $\mathfrak{A} \models^- B$

To state the truth/falsity conditions for quantified statements, we assume that the language has been augmented with a new constant, k_d , for every $d \in D$, such that $\delta(k_d) = d$.⁵ We then have:

- $\mathfrak{A} \models^+ \forall x A$ iff for all $d \in D$ $\mathfrak{A} \models^+ A_x(k_d)$
- $\mathfrak{A} \models^- \forall x A$ iff for some $d \in D$ $\mathfrak{A} \models^- A_x(k_d)$
- $\mathfrak{A} \models^+ \exists x A$ iff for some $d \in D$ $\mathfrak{A} \models^+ A_x(k_d)$
- $\mathfrak{A} \models^- \exists x A$ iff for all $d \in D$ $\mathfrak{A} \models^- A_x(k_d)$

For closed formulas, A , and sets of closed formulas, Σ :

- \mathfrak{A} is a *model* of A iff $\mathfrak{A} \models^+ A$
- \mathfrak{A} is a *model* of Σ iff $\mathfrak{A} \models^+ A$, for every $A \in \Sigma$
- $\Sigma \models_{LP} A$ iff every model of Σ is a model of A

Note that all of this is *exactly* the same as classical logic (CL), except that the semantics of LP allow truth and falsity to overlap.

Next, some simple properties of LP . Call an interpretation *classical* if $\delta^+(P_n) \cap \delta^-(P_n) = \emptyset$, for every P_n . Then clearly, every classical interpretation is an LP interpretation. Hence:

- $\Sigma \models_{LP} A \Rightarrow \Sigma \models_{CL} A$

However:

- $\{A, \neg A\} \not\models_{LP} B$

Hence:

- $\Sigma \models_{CL} A \not\Rightarrow \Sigma \models_{LP} A$

But:⁶

- $\emptyset \models_{CL} A \Leftrightarrow \emptyset \models_{LP} A$

⁵ This is not essential. It is just a simple way of avoiding talk of satisfaction.

⁶ For the proof, see Priest (2006a), ch. 5.

2.2 Collapsed Models. We now come to an important piece of *LP* metatheory: the Collapsing Lemma.

Let \mathfrak{A} be any *LP* interpretation (and so any classical interpretation), and let \sim be any equivalence relation on the domain of \mathfrak{A} , which is also a congruence relation on the interpretations of the function symbols in the language. That is:

- $d_1 \sim e_1, \dots, d_n \sim e_n \Rightarrow \delta(f_n)(d_1, \dots, d_n) \sim \delta(f_n)(e_1, \dots, e_n)$

If $d \in D$, let $[d]$ be the equivalence class of d .

Now define the *collapsed interpretation*, $\tilde{\mathfrak{A}} = \langle \tilde{D}, \tilde{\delta} \rangle$, as follows:

- $\tilde{D} = \{[d] : d \in D\}$
- For every constant, c , $\tilde{\delta}(c) = [c]$
- For every n -place function symbol, f_n , $\tilde{\delta}(f_n)([d_1], \dots, [d_n]) = [\delta(f_n)(d_1, \dots, d_n)]$
- For every n -place predicate, P_n :
 - $\langle [d_1], \dots, [d_n] \rangle \in \tilde{\delta}^+(P_n)$ iff for some $e_1 \sim d_1, \dots, e_n \sim d_n$ $\langle e_1, \dots, e_n \rangle \in \delta^+(P_n)$
 - $\langle [d_1], \dots, [d_n] \rangle \in \tilde{\delta}^-(P_n)$ iff for some $e_1 \sim d_1, \dots, e_n \sim d_n$ $\langle e_1, \dots, e_n \rangle \in \delta^-(P_n)$

It is easy to check that the collapsed interpretation is indeed an *LP* interpretation. In effect, the collapse identifies all the members of an equivalence class, to deliver an object which has all the properties of the things in it.

One may now establish the Collapsing Lemma:

If \mathfrak{A} is any *LP* interpretation, and $\tilde{\mathfrak{A}}$ is any collapse then:

- $\tilde{\delta}(t) = [\delta(t)]$
- If $\mathfrak{A} \Vdash^+ A$ then $\tilde{\mathfrak{A}} \Vdash^+ A$
- If $\mathfrak{A} \Vdash^- A$ then $\tilde{\mathfrak{A}} \Vdash^- A$

In other words, when an interpretation is collapsed, truth and falsity values are preserved (though other things may become true or false in the process). The proof of the Lemma is by the appropriate inductions, and I omit it here.⁷

3. Inconsistent Arithmetic

3.1 Inconsistent Models. We can now turn to Gödel's theorem. To do so, we need to apply the Collapsing Lemma to establish some facts about models of inconsistent arithmetic.⁸

In what follows, let the language be the standard one for first-order arithmetic (with one constant, 0, and function symbols for successor, addition, and multiplication). Let \mathbf{N} be the set of sentences in this language true in the standard (classical) model of arithmetic. Let \mathfrak{M} be any model of \mathbf{N} .

If $\tilde{\mathfrak{M}}$ is any collapsed interpretation, $\tilde{\mathfrak{M}}$ is a model of \mathbf{N} , by the Collapsing Lemma. Let me give a couple of examples of such collapsed models.

Example 1

- \mathfrak{M} is any classical non-standard model of arithmetic

⁷ Details can be found in Priest (2006a), p. 229 f.

⁸ For this, and much further discussion, see Priest (2006a), ch. 17, and the references cited therein.

- $d \sim e$ iff (d and e are standard and $d = e$) or (d and e are non-standard)

What \sim does, in effect, is to leave all the standard numbers alone, but identify all the non-standard numbers. It is easy to check that \sim is an equivalence relation, and a congruence with respect to the function symbols. We may draw the diagram of the successor relation in the collapsed model as follows:

$$0 \rightarrow 1 \rightarrow 2 \rightarrow \dots \rightarrow i \rightarrow$$

As is clear, $\tilde{\mathfrak{M}} \models \exists x x = x'$, but $\tilde{\mathfrak{M}} \not\models 0 = 0'$. Hence, $\tilde{\mathfrak{M}}$ is a model of \mathbf{N} , and the set of things true in it is inconsistent but non-trivial (i.e., not the set of all formulas).

Example 2

- \mathfrak{M} is the standard model of arithmetic
- n and p are a natural numbers greater than 1.
- $d \sim e$ iff ($d, e < n$ and $d = e$) or ($d, e \geq n$ and $d = e \pmod{p}$)

This relation leaves numbers less than n alone, but numbers greater than or equal to n behave as in arithmetic mod p . Again, it is easy to check that this is an equivalence relation, and a congruence relation on the function symbols. We may depict the collapsed interpretation as follows.

$$\begin{array}{ccccccc} & & & & n+p-1 & \leftarrow & \dots & \leftarrow & n+3 \\ & & & & \downarrow & & & & \uparrow \\ 0 & \rightarrow & 1 & \rightarrow & \dots & \rightarrow & n & \rightarrow & n+1 & \rightarrow & n+2 \end{array}$$

As is clear, $\tilde{\mathfrak{M}} \models \exists x x = x' \dots'$ (with p occurrences of the successor function symbol), and $\tilde{\mathfrak{M}} \not\models 0 = 0'$. Hence, again, $\tilde{\mathfrak{M}}$ is a model of an inconsistent but non-trivial set of sentences that contains all the sentences true in the standard model. Moreover, in this case, $\tilde{\mathfrak{M}}$ is finite. Hence, the set of sentences true in $\tilde{\mathfrak{M}}$ is decidable. (The particular and universal quantifiers behave like finite disjunctions and conjunctions, respectively.) Hence, it is axiomatizable.

We see, then, that there are non-trivial but inconsistent axiomatic *LP* theories which contain all the truths in the standard model. There are, therefore axiomatic *LP* theories which are complete but inconsistent.⁹ Thus, in the disjunctive statement of Gödel's Theorem stated in 1.2, one can no longer ignore the inconsistency disjunct.

3.2 Back to the Gödel Sentence. Let us now turn to look at the behaviour of the “Gödel sentence” in these inconsistent theories.

Let T be any axiomatic arithmetic such that $T \supseteq \mathbf{N}$. Let $Prov x$ be the arithmetic predicate which defines T in the standard model. That is:

- If $T \vdash A$ then $Prov \langle A \rangle \in \mathbf{N}$, so $T \vdash Prov \langle A \rangle$
- If $T \not\vdash A$ then $\neg Prov \langle A \rangle \in \mathbf{N}$, so $T \vdash \neg Prov \langle A \rangle$

⁹ I note that in the case of Example 2, the model is finite. But this is unnecessary. There are infinite models of the same kind, i.e., whose theories are axiomatic. This is proved in Paris and Sirokofskich (2008).

Let G be $\neg Prov\langle G\rangle$. Then either $T \vdash G$ or $T \not\vdash G$. In either case, $T \vdash \neg Prov\langle G\rangle$. So $T \vdash G$, and $T \vdash Prov\langle G\rangle$. That is, the Gödel sentence and its negation are both provable—as opposed to the consistent case, where neither is provable.¹⁰

3.3 Some Philosophical Issues. So much for purely technical matters. Let me now make a few philosophical observations.

Behind Gödel’s proof of the Incompleteness Theorem—and of course, there are others—there is a recognisable paradox concerning the sentence ‘this sentence is not provable’, which is clearly in the same family as the Liar. We may call this *Gödel’s Paradox*, and it goes as follows. Let G be $\neg Prov\langle A\rangle$.

- Suppose that $Prov\langle G\rangle$
- Then G (what is provable is true)
- That is, $\neg Prov\langle G\rangle$
- Hence, $\neg Prov\langle G\rangle$ (is proved)
- That is, G
- So $Prov\langle G\rangle$ (we have just proved it)

Now, of course, this argument cannot be formalised in the arithmetic if it is consistent: the second line cannot be enforced, however natural it is. However, in a dialethic solution to the Liar Paradox, the thought is that there is nothing wrong with the reasoning which produces contradiction. It is just fine, and we should accept the contradiction it delivers (though the use of a paraconsistent logic prevents an arbitrary conclusion following). If this is correct then, it would seem, we should accept the same conclusion in the case of the Gödel’s Paradox. The inconsistent arithmetic, then, gets things exactly right.

If one asks why one cannot push through the paradoxical argument in the case of consistent arithmetics, the answer is that Löb’s Theorem gets in the way. For the theorem tells us that, for any A , if $\vdash Prov\langle A\rangle \supset A$ then $\vdash A$. Hence, we cannot infer A from $Prov\langle A\rangle$, for arbitrary A .

Now, in truth, there is already something strange about Löb’s Theorem. Given that the axiom system is sound, of course one ought to be able to infer A from the fact that A is proved. Indeed, for any A , $Prov\langle A\rangle \supset A$ is true in the standard model!¹¹ Intuitively, one should expect all the Löb sentences to hold in the correct arithmetic. And they do in the inconsistent arithmetics. To see this, note that for any A , either $T \vdash A$ or $T \not\vdash A$. In the first case, obviously $T \vdash Prov\langle A\rangle \supset A$. In the second case, $T \vdash \neg Prov\langle A\rangle$, so $T \vdash Prov\langle A\rangle \supset A$.

¹⁰ It might be wondered what happens to Gödel’s Second Theorem in this context. This theorem states that given a consistent axiomatic arithmetic of the kind we are dealing with, the canonical statement of its consistency cannot be proved in the theory. In the case of inconsistent arithmetics, since they are inconsistent, one would not be expect the theory to be able to prove its consistency. However, if the theory is non-trivial, one might hope that the theory can prove its own non-triviality. Indeed in the case of the inconsistent axiomatic arithmetics we are considering, it can. For take some unprovable sentence, A . Then since $T \not\vdash A$, $\neg Prov\langle A\rangle \in \mathbf{N}$, and so $T \vdash \neg Prov\langle A\rangle$. (Warning: in some theories, we might be able to prove $Prov\langle A\rangle$ as well. See, further, Priest (2006a), 17.4.)

¹¹ In case anyone is puzzled by this technically, note that one can add all instances of the Löb sentence to obtain a consistent theory. The cost is that $Prov$ no longer represents provability in this new system.

4. Objections

Let me end with a brief mention of a few philosophical objections one might be tempted to make at this point.

Perhaps the first thought would be that one cannot prove enough in the inconsistent arithmetics. After all, *modus ponens* for \supset is not valid. That thought can be set aside quickly, though. Everything true in the standard model is provable in the theory: everything one could want.

But the next obvious thought is that one can prove too much. True, we have everything true in the standard model; but we have some of their negations too. Why, however, should such contradictions be a problem? One reason one might give is that, by the Principle of Non-Contradiction, contradictions are not true, so one should not want them to be provable. Clearly, the objection turns on the correctness of the Principle of Non-Contradiction, and so now the debate must turn to the question of what reason one has for supposing this to be true. This is not the place to go into the matter. Let me just say that arguments for the unrestricted validity of the Principle are exceptionally hard to find.¹²

The next thought is a more subtle one. We apply arithmetic in science for many purposes. Contradictions in arithmetic would ruin these applications. This does not follow, however. Thus, suppose that the inconsistent arithmetic is given as in Example 2. Then all the equations for numbers less n behave consistently.¹³ So any sentence where the quantifiers are bounded by a number less than n also behave consistently. Now, for practical purposes, we are rarely concerned with *really* large numbers. Thus suppose that there is an inconsistent natural number, and so a least, $m + 1$. Suppose that this is so large that it has no empirical meaning. (Maybe m is $A(n, n)$, where A is the Ackermann function, and n is the number of elementary particles in the cosmos.) Any sentence with quantifiers bounded by m is consistent. So the contradictions have no practical implications.¹⁴

5. Conclusion

Whatever one makes of these matters, it is clear that applying a paraconsistent logic to arithmetic opens up a whole new dimension on matters concerning Gödel's Incompleteness Theorem. Indeed, allowing for the possibility of inconsistency-tolerant theories, does this in many areas. But Gödel's Theorem is undoubtedly one of the most profound pieces of mathematics in the last 100 years. The new dimensions concerning this are, then, equally profound.

¹² For just the start of the debate, see Priest (1989). See also the essays in Priest (200b).

¹³ And of course, in Example 1, all the natural numbers behave consistently.

¹⁴ For some further discussion, see Priest (1994).

Bibliography

- [1] Paris, J. B., and Sirokofskich, A. (2008), ‘On *LP* Models of Arithmetic’, *Journal of Symbolic Logic* 73: 212-226.
- [2] Priest, G. (1994), ‘What could the Least Inconsistent Number be?’, *Logique et Analyse* 37: 3-12.
- [3] Priest, G. (1998), ‘What’s so Bad about Contradictions?’, *Journal of Philosophy* 95: 410–26; reprinted as ch. 1 of Priest, G., Beall, J. C., and Armour-Garb, B. (eds.), *New Essays on the Law of Non-Contradiction*, Oxford: Oxford University Press, 2004.
- [4] Priest, G. (2006a), *In Contradiction*, 2nd edn, Oxford: Oxford University Press.
- [5] Priest, (2006b), *Doubt Truth to be a Liar*, Oxford: Oxford University Press.
- [6] Priest, G. (2008), *Introduction to Non-Classical Logic*, 2nd edn, Cambridge: Cambridge University Press.