Consistency, Paraconsistency, and the Logical Limitative Theorems

Graham Priest

1 Introduction

Many sorts of thing may be subject to limitation: how fast something can travel; what one can say legally; what can be imagined. This talk concerns just one of these many: logic.[1] As is familiar to any student of modern logic, there are certain things that cannot be done within logical confines. These are the limitative theorems of classical metatheory. A list of the most familiar would include the following, which can be broken down into two sub-groups. (The statements are rather rough and ready, but will do for the present.)

Limitative Theorems of Metalogic

1. Church's Theorem: first-order logic cannot be decided by an algorithm.
2. Löwenheim-Skolem Theorems: no infinite structure can be characterised by a first-order theory.

Limitative Theorems of Metamathematics

1. Tarski's Theorem: no theory can contain a truth predicate for its own language.
2. Gödel's first Incompleteness Theorem: no axiomatisable arithmetic can be complete.

[1]Specifically, first-order logic. Second-order logic raises a number of interesting questions concerning limitations too; I will not consider these here.

3. Gödel's second Incompleteness Theorem: no theory of arithmetic can prove its own consistency.

When limitation is at issue, it always makes sense to ask whether the limitation is absolute or relative. For example, if a body is accelerating in a vacuum, then, if the Special Theory of Relativity is right, it can never surpass the speed of light. And this limitation is absolute: there is nothing that can be done to take the body beyond this limit. Now consider weight-lifting. There is a maximum weight that a person can lift, say in the Olympic Games. We do not know what it is, but basic bio-physics assures us that there must be such a limit. Yet if a person is allowed to take steroids or other body-building drugs, they will certainly be able to lift more. The limit to what can be lifted is therefore relative to drug-consumption.

Now what of the limitations of logic? Are these absolute or relative? It is natural to suppose that they are absolute. If the bounds of logic are not absolute, what could be?! Yet, as one reflects, one may start to wonder about this. In particular, most of the classical limitative results involve consistency in one way or another—if only in that many of them are standardly proved by *reductio*. Now what if we may go *beyond consistency*—into the transconsistent. Are the limitations then removed? That is the subject of this talk. The answer to this question is complex and nuanced, and a full one cannot be given in one lecture. What can be done, and what I will do here, is give the outlines of an answer. Much of the talk simply surveys what is known about the subject; but the final section strays from this safe ground, to areas that are highly speculative.

2 Background on Paraconsistency

Of course, the question of whether the limitations of logic are relative to consistency makes sense only if the idea that one can go beyond the consistent itself makes sense. For much of the last century—if not for many other centuries in the West—it would have been supposed that it does not. To be inconsistent is to be incoherent. This view has been challenged, and in my view conclusively disposed of (to the extent that anything can be in philosophy), by work in formal logic in the last 30 years. The development of paraconsistent logics, that is, logics in which contradictions do not entail everything, has shown that inconsistency is not to be equated with incoherence. Inconsistent theories can have a quite determinate and non-trivial structure.

To understand the discussion that follows one needs to have some grasp of what paraconsistent logics are like. There are, in fact, many such logics.[2] Moreover, the issue of what happens to the limitative theorems may depend on exactly which paraconsistent logic is deployed. However, it is not my aim here to survey all of the various possibilities; it is to show the sort of thing that may happen once one deploys a paraconsistent logic. So let me start by explaining just one paraconsistent logic. (In point of fact, much of the discussion carries over to other paraconsistent logics, without too much change.) This is the logic LP. I choose this logic partly because I am fond of it,[3] but more importantly because it is one of the simplest and most natural paraconsistent logics. In subtle discussions of the kind that we will be engaged in, it always helps to keep the framework simple where possible.

Take for the language that of standard first-order logic—function symbols are optional; to keep things simple, we suppose a countable vocabulary and eschew free variables. An interpretation, I, is a pair, $\langle D, d \rangle$, where D is a non-empty domain (of quantification) and:

- for every constant, c, $d(c) \in D$
- for every n-place predicate, P, $d(P) = \langle d^+(P), d^-(P) \rangle$, where $d^+(P) \cup d^-(P) = D^n$
- for every n-place function symbol, f, $d(f) : D^n \to D$
- $d^+(=) = \{\langle x, x \rangle : x \in D\}$

(Note that for any P, including identity, $d^+(P)$ and $d^-(P)$ do not have to be disjoint.) d is extended so that it assigns a denotation to every term by recursion in the usual way:

- $d(ft_1...t_n) = d(f)(d(t_1), ..., d(t_n))$

If A is any formula, it may be true or false (or both) in the interpretation I. We write these as $I \Vdash_t A$ and $I \Vdash_f A$, respectively, and the notions may be defined by a joint recursion as follows:

- $I \Vdash_t Pt_1...t_n$ iff $\langle d(t_1), ..., d(t_n) \rangle \in d^+(P)$

[2] For a survey, see Priest (200a). Most of the proofs of the results referred to below can be found there.

[3] See Priest (1987).

- $I \Vdash_f Pt_1...t_n$ iff $\langle d(t_1), ..., d(t_n)\rangle \in d^-(P)$
- $I \Vdash_t \neg A$ iff $I \Vdash_f A$
- $I \Vdash_f \neg A$ iff $I \Vdash_t A$
- $I \Vdash_t A \vee B$ iff $I \Vdash_t A$ or $I \Vdash_t B$
- $I \Vdash_f A \vee B$ iff $I \Vdash_f A$ and $I \Vdash_f B$
- $I \Vdash_t A \wedge B$ iff $I \Vdash_t A$ and $I \Vdash_t B$
- $I \Vdash_f A \wedge B$ iff $I \Vdash_f A$ or $I \Vdash_f B$
- $I \Vdash_t \exists x A(x)$ iff $I \Vdash_t A(c)$ for some constant, c
- $I \Vdash_f \exists x A(x)$ iff $I \Vdash_f A(c)$ for all constants, c
- $I \Vdash_t \forall x A(x)$ iff $I \Vdash_t A(c)$ for all constants, c
- $I \Vdash_f \forall x A(x)$ iff $I \Vdash_f A(c)$ for some constant, c

In the truth/falsity conditions for the quantifiers, we assume, to keep things simple, that the language has been augmented by constants such that every member of D is named by one of them, if this was not already the case. Validity is defined in the way that one might expect. Say that I is a *model* of A iff $I \Vdash_t A$, and a model of a set of sentences, Σ, iff I is a model of every member of Σ. Then:

- $\Sigma \vDash A$ iff every model of Σ is a model of A

Some comments about $\vDash$. First, note that the interpretations of classical logic are, in effect, special cases of LP interpretations, namely those in which for every P, $d^+(P) \cap d^-(P) = \phi$. Hence, $\vDash$ is a sub-relation of classical consequence. It is, however, a proper sub-relation: it is not difficult to check that it is paraconsistent, that is, there are A and B such that $\{A, \neg A\} \nvDash B$. Though the consequence relation of LP is different from that of classical logic, it is not difficult to prove that the logical truths of the two logics are the same. The logic can also be furnished with a sound and complete proof-system. The compactness theorem follows from this in the usual way. Finally, note that the language has no conditional connective. We may define $A \supset B$ as $\neg A \vee B$, as is standard in classical logic, but, as is not difficult to check, *modus ponens* for such a connective fails. One can augment the language and the semantics with a ponible conditional in various ways, but for present purposes this is unnecessary. The aim of the present enterprise is to see what may happen to the classical limitative results in a paraconsistent context, and classical theories can all be expressed in the above vocabulary.

While we are on technical preliminaries, let us get one more important one out of the way. This is the *Collapsing Lemma.* Let I be any interpretation with domain D. Let $\sim$ be any equivalence relation on D; write the equivalence class of x as $[x]$. If there are any function symbols in the language, we also require $\sim$ to be a congruence relation on their denotations in I. We now define a collapsed interpretation, $I_\sim$. In effect, this simply identifies all the members of an equivalence class into a single individual which possesses all the properties of its members (even when these are contradictory). The formal definition is as follows. $I_\sim = \langle D_\sim, d_\sim \rangle$, where:

- $D_\sim = \{[x] : x \in D\}$
- $d_\sim(c) = [d(c)]$
- $d_\sim(f)([x_1], ..., [x_n]) = [d(f)(x_1, ..., x_n)]$
- $\langle [x_1], ..., [x_n] \rangle \in d^+_\sim(P)$ iff $\exists y_1 \in [x_1], ..., \exists y_n \in [x_n]$, $\langle y_1, ..., y_n \rangle \in d^+(P)$
- $\langle [x_1], ..., [x_n] \rangle \in d^-_\sim(P)$ iff $\exists y_1 \in [x_1], ..., \exists y_n \in [x_n]$, $\langle y_1, ..., y_n \rangle \in d^-(P)$

It is not difficult to check that $I_\sim$ is an LP interpretation. The Collapsing Lemma, which can now be proved by an induction, tells us that for all A:

- if $I \Vdash_t A$ then $I_\sim \vDash_t A$
- if $I \Vdash_f A$ then $I_\sim \vDash_f A$

In other words, truth values are not lost in the collapse. In particular, therefore, if I is a model of some set of sentences, so is $I_\sim$.

3 Limitative Theorems of Metalogic

The foregoing material is sufficient for us to survey what happens to the classical limitative results in a paraconsistent context. Let us start with the results of metalogic.

Church's Theorem Propositional LP is decidable by simple truth-table techniques. However, first-order LP shares its logical truths with classical logic. Hence, it is not decidable (even without identity). Church's theorem therefore stands.

Löwenheim-Skolem Theorems The Löwenheim-Skolem Theorems hold for LP. If identity does not occur in the language, they hold in a very strong form indeed: every theory has a model of every cardinality! Consider a trivial model of cardinality κ. This has domain of size κ, and for every n-place predicate P, $d^+(P) = d^-(P) = D^n$. It is not difficult to show that this is a model of every theory.

What happens once identity is included in the language? A standard proof of the upwards Löwenheim-Skolem depends on an appropriate version of the compactness theorem. This holds for LP, and can be applied in much the usual way to show that if a set of sentences has a model of infinite cardinality κ it has a model of every cardinal greater than κ. A standard proof of the downwards Löwenheim-Skolem Theorem goes via the completeness theorem. A straightforward modification of this proof shows that if a set of sentences has an infinite LP model, it has a countable model (and so a model of all infinite cardinalities). However, something even stronger holds in LP. If a set of sentences (with no function symbols) has a model of any cardinality, it has a model of every lower cardinality! This follows from the Collapsing Lemma. Just take any model of cardinality κ. Consider any equivalence relation on its domain that generates λ equivalence classes, $1 \leq \lambda \leq \kappa$. The Collapsing Lemma tells us that the collapsed model which this equivalence relation generates is a model of the original set of sentences, and if $\lambda > 1$ the model is non-trivial (since some identities are not true in the model).

What we have seen, to summarise, is that all the limitative theorems of classical metalogic carry over to a paraconsistent context. Indeed, because of the Collapsing Lemma, they apply in *even stronger* forms. The situation is different when we turn from metalogic to metamathematics, which we now do.

4 Limitative Theorems of Metamathematics

Let us take a first-order mathematical theory. Any appropriately strong mathematical theory will do. For the sake of definiteness, let us take arithmetic. The language is the usual one. Successor, addition, and multiplication may be represented by appropriate function symbols, $'$, $+$, and $\times$, or by appropriate predicates. Let $\mathbf{m}$ be the numeral of the number m. We suppose that we have a gödel coding for the language. If A is a formula, let $\langle A \rangle$ be the numeral of its code number. This gives us the machinery to represent self-reference. By appropriate coding, for any formula with one free variable, $A(x)$, we can construct a formula, $A(\mathbf{n})$, with code number n. Finally, let Σ be some theory in this language which can define all recursive sets (i.e., sets whose characteristic functions are recursive). In particular, for every such set, X, there is a formula, $A(x)$, such that:

if $m \in X$ then $\vdash_\Sigma A(\mathbf{m})$

if $m \notin X$ then $\vdash_\Sigma \neg A(\mathbf{m})$

where $\vdash_\Sigma$ is deducibility within Σ.

Tarski's Theorem Now, suppose that there were a formula of one free variable, $T(x)$, such that:

Tr $A \vdash_\Sigma T(\langle A \rangle)$

Tl $T(\langle A \rangle) \vdash_\Sigma A$

If there were such a thing, we could find a formula, A, with code n, of the form $\neg T(\mathbf{n})$. For such an A, by **Tr** and **Tl**, A and $\neg A$ are inter-deducible in Σ, and hence Σ is inconsistent. Thus, no consistent theory can represent its own truth predicate. But if Σ may be inconsistent, this is obviously no problem. Indeed, we can actually construct inconsistent but non-trivial theories where truth can be represented in this way. Whether T is itself a formula expressible in the language of arithmetic depends on Σ. But such a predicate can always be added to the language (with a suitable extension of the gödel coding, etc.).

In particular, let Σ be the set of truths in the standard model of arithmetic. Let T be a new predicate. Then the addition of the two rules of inference **Tr** and **Tl** to Σ gives a theory that is inconsistent but non-trivial. In this theory no purely arithmetic formula is (provably) inconsistent; indeed, the formulas that are are all ungrounded, in the sense of Kripke. The proof is not too difficult, but it is too long to go in to here. We have seen enough to note that in a paraconsistent context Tarski's theorem fails. There are inconsistent but non-trivial theories that contain their own truth predicates.

Let us move on to Gödel's theorems. To discuss these, we apply the Collapsing Lemma again. Let N be the set of sentences of first-order arithmetic true in the standard model. Let I be any model of N. Let $\sim$ be an equivalence relation with a finite number of equivalence classes ≥ 2—and if successor, addition and multiplication are represented by function symbols, let $\sim$ also be a congruence with respect to these. For example, and for the sake of definiteness in what follows, I might be the standard model itself, and $\sim$ the relation:

- $x \sim y$ iff $x, y \geq j$ or $(x, y < j$ and $x = y)$

for some $j \geq 2$. If we apply the collapse to I using this relation, we obtain a finite (non-trivial) model, J, and the Collapsing Lemma tells us that J is a model of N.

Now, J is a finite model, and so whether or not something holds in it can be determined in an effective way. (Quantified sentences are, effectively, finite conjunctions and disjunctions.) Hence its membership is recursive. In other words, if Σ_J is the set of sentences true in J, Σ_J is decidable. Thus, the proof predicate of Σ_J is definable in N. That is, there is a predicate of one free variable, $P(x)$, such that:

if $A \in \Sigma_J$ then $I \Vdash_t P\langle A \rangle$

if $A \notin \Sigma_J$ then $I \Vdash_t \neg P\langle A \rangle$

(I write $P\langle A \rangle$ rather than the more cumbersome $P(\langle A \rangle)$.) By the Collapsing Lemma, this holds of J, too. Thus:

P1 if $A \in \Sigma_J$ then $\vdash_{\Sigma_J} P\langle A \rangle$

P2 if $A \notin \Sigma_J$ then $\vdash_{\Sigma_J} \neg P\langle A \rangle$

Gödel's First Incompleteness Theorem One way to state the first Incompleteness Theorem is that any complete axiomatisable theory of arithmetic is inconsistent—where a complete theory here is one that proves, for every A, either it or its negation. Paraconsistency does not challenge this result. It is usually taken to follow from this that any complete axiomatisable theory of arithmetic is trivial. This consequence can now be seen to be false. Σ_J is decidable, and *a fortiori* axiomatic; it contains N, and *a fortiori* is complete; it is inconsistent, but non-trivial.

What about the famous undecidable sentence? This is a sentence, A, of the form $\neg P\langle A\rangle$. If $A \in \Sigma_J$ then $P\langle A\rangle \in \Sigma_J$, by **P1**. If, on the other hand, $A \notin \Sigma_J$ then $\neg P\langle A\rangle \in \Sigma_J$, by **P2**. But since $A \notin \Sigma_J$, $\neg A \in \Sigma_J$; i.e., $\neg\neg P\langle A\rangle \in \Sigma_J$ (and so $P\langle A\rangle \in \Sigma_J$). In either case, $P\langle A\rangle \wedge \neg P\langle A\rangle$ holds in Σ_J. Thus the "undecidable" sentence, as one might expect, turns out to be decidable, but inconsistently so. In this, it just mirrors the intuitive paradox concerning the sentence 'This sentence is unprovable'. Suppose this is provable; then it is unprovable. Hence it is unprovable. Thus, it is true, and we have just proved this. The sentence is both provable and unprovable. We might call this paradox *Gödel's paradox*. It is a version of the Knower paradox.

Gödel's Second Incompleteness Theorem A standard formulation of the second Incompleteness Theorem is to the effect that a consistent theory cannot prove its own consistency. Since we are now dealing with inconsistent theories, nothing we have done so far affects this.[4] However, classically, inconsistency and triviality go together. So a classically equivalent statement is to the effect that no non-trivial theory can prove its own non-triviality. The above construction shows this to be false. Thus, the sentence $\mathbf{0} = \mathbf{1}$ is not in Σ_J. Hence, by **P2**, $\neg P\langle \mathbf{0} = \mathbf{1}\rangle$ is also in Σ_J. (Note, however, that this does not rule out its negation also being in Σ_J.)

In summary, then, there are certainly ways of stating the limitative results of classical metamathematics that survive the transition to paraconsistency—specifically ones where the consistency clause is made explicit. However, the absolute impossibilities in question fail once we take seriously the possibility of inconsistent theories. In particular, there are inconsistent but non-trivial theories which allow all the things ruled out classically.

5 Decidability Revisited: Inconsistent Computation

So far, we have been concerned with the investigation of theories that may be inconsistent, theories based on a paraconsistent logic. What of the theory/logic in which our results about these theories are established? As usual

[4]Though, it should be noted, there are consistent theories of arithmetic based on a relevant logic—one kind of paraconsistent logic—that can be proved to be consistent in the theory itself. This insufficiently appreciated result was established by Meyer (1978).

in the subject, this has been left at a purely informal level. A standard assumption is that the proofs can be regimented in a formal theory with an underlying classical logic, such as Zermelo-Fraenkel set theory. And presumably they can. But suppose that our metatheory is itself inconsistent. This raises a number of important questions—not least, what an inconsistent metatheory should be like. This is not the place to go into these questions.[5] I merely want to illustrate the sort of thing that may happen if the metatheory itself is allowed to be inconsistent.

Let us suppose that we employ in our metatheory an inconsistent arithmetic. Suppose, indeed, that the set of truths of arithmetic is actually inconsistent. This is obviously a much stronger assumption than that a paraconsistent logic provides the correct canons of validity. Such a logic allows for the possibility of inconsistent but non-trivial theories, but this, in itself, does not entail that any of these inconsistent theories is actually true (*dialetheism*).

The assumption that true arithmetic is inconsistent is not, in fact, as outrageous as it may at first seem. We have already seen that there is a paradox associated with the notion of proof. It is not implausible that in the true arithmetic this generates inconsistency.[6] This does not mean that we have to reject things that hold in the "standard model" of arithmetic. We may suppose that all these things are true. It is just that this theory is not complete—in the sense that more things will be true as well. We have already seen that there are inconsistent theories of arithmetic that contain the set of things true in the standard model—and that also encode Gödel's paradox. Not all these theories are decidable, but we have seen that some of them are.[7] Let us make the further assumption that true arithmetic is one such theory—for example, Σ_J of the previous section.

This assumption has consequences that reflect back on some of the limitative results themselves. For example, consider the matter of the decidability of first-order logic (both classical and paraconsistent). Provability in this theory can be expressed by a Σ_1 sentence. That is, there is a formula of first-order arithmetic $A(x, y)$ which contains no quantifiers such that for any formula, B, B is provable iff $\exists x A(x, \langle B \rangle)$ is true. But now, if arithmetic is decidable, the truth (and/or falsity) of this sentence can be decided by an algorithm. (In the case of Σ_J, this is little more than a method for com-

[5]The shape of an inconsistent set-theoretic metatheory that that has much of the power of ZF is discussed in Priest (200b).

[6]For further discussion, see Priest (1987), ch. 3 and Priest (1994).

[7]For the general structure of inconsistent arithmetics, see Priest (1997) and (2000).

puting the values of identities and their negation, plus truth tables.) Thus, first-order logic is itself decidable.

In the same way, any problem that can be represented by a sentence of first-order arithmetic is decidable. Thus, as another example, consider the halting problem. Consider an algorithm/computer program with code e, and some input i. The claim that that program with that input terminates can be expressed by a sentence of first-order arithmetic, $\exists x T_1(x, \mathbf{e}, \mathbf{i})$, where $T_1(x, y, z)$ is the one-input Kleene T-predicate. Thus the halting problem is decidable!

Of course, we must be prepared for the appearance of strange things in our new landscape. Thus, consider the sentence of first-order arithmetic that states the Halting Theorem:

$$\neg\exists y \forall e \forall i \exists x (T_1(x, y, [e, i]) \wedge \forall z (U(x, z) = \mathbf{0} \equiv \exists w T_1(w, e, i)))$$

where $U(x, z)$ means that z is the output of computation with code x, and $[e, i]$ is the code of the ordered pair of e and i. Since this holds in the standard model, it is provable in the arithmetic. Is it the case that there both is and is not an algorithm that decides whether a computation halts? Worse, this inconsistency is presumably due to the inconsistent behaviour of the predicate T_1. For the fact that a computation terminates, so that, for some k, $T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ holds, does not rule out the fact that it does not terminate, $\neg\exists x T_1(x, \mathbf{e}, \mathbf{i})$, so that $\neg T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ as well. But how can this be? As Shapiro puts it, in a slightly different context:[8]

> On all accounts ... we have that k is the code of a computation of program e with input i. This can be verified with a painstaking, but completely effective check. How can the dialetheist maintain that k is not the code of such a computation. What does it mean to say this? Since $\neg T_1$ is a recursive predicate, we can supposedly verify—at the same time, in almost exactly the same

[8] Shapiro (200a), p. 14 of ms. Let me comment on a minor error in Shapiro's paper. He says (p. 15) that if this contradiction were true, it would be derivable in Peano Arithmetic (PA) on the ground that all true Δ_0 statements are provable in PA. It is not difficult to see that the proof of the latter fact breaks down in the present context. The proof is by recursion. Consider the basis case, which concerns identity. Suppose, e.g., that $\mathbf{0} = \mathbf{n}$ is true. Then (classically), n must be 0, and we know that $\mathbf{0} = \mathbf{0}$ is provable in PA. But in Σ_j for example, n may be j (> 0), and $\mathbf{0} = \mathbf{j}$ is not provable in PA. Indeed, if the true Δ_0 sentences are inconsistent, and PA is consistent, then it is precisely not the case that all true Δ_0 sentences are provable in PA.

> way—that k is not the code of a computation of program e with input i. How? I must admit that I cannot make anything of this supposed possibility.

In the passage in question, Shapiro is, in fact, talking of codes of proofs rather than computations, but the considerations are exactly the same. A computation that stops and goes on for ever is the same as a proof tableau that finishes but that goes on forever. I have taken the liberty of modifying Shapiro's quote accordingly.

Several points are relevant here. First, Shapiro has considerably overstated the case. In a paraconsistent logic, and particularly in the inconsistent models, truth and falsity (i.e., truth of negation) fall apart. If we are dealing with a decidable theory, different procedures are therefore necessary to determine the truth of a sentence and its negation. For example, in Σ_J, to determine the truth of the equation $\mathbf{m} = \mathbf{n}$, we compute to see whether m and n are less than j; we set the equation to *true* if they are, or if they are not and m and n are the same. To determine the falsity of the equation $\mathbf{m} = \mathbf{n}$, we compute to see whether m and n are less than j; we set the equation to *false* if they are, or if they are not and m and n are distinct. Hence, there is no problem about how the algorithm can determine both $T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ and $\neg T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ to be true.

It remains the case that they both are true. But how can k both be the code of a computation and not be the code? Again, one must recall that truth and falsity have fallen apart. $T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ records the truth of ks being the appropriate code. $\neg T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ does *not* record its untruth, but the truth of ks being *distinct from* the appropriate code. Let me illustrate. Suppose that the code is 36, then $T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ says, in effect, that $\mathbf{k} = \mathbf{36}$. Similarly, $T_1(\mathbf{k}, \mathbf{e}, \mathbf{i})$ says, in effect, that $\neg \mathbf{k} = \mathbf{36}$. But this latter sentence does not rule out k's being 36: this holds *as well*.

In the same way, though $\exists x T_1(x, \mathbf{e}, \mathbf{i})$ may record the existence of a code of a terminating computation, $\neg\exists x T_1(x, \mathbf{e}, \mathbf{i})$ does not rule out its existence. It is equivalent to $\forall x \neg T_1(x, \mathbf{e}, \mathbf{i})$, i.e., $\neg x = \mathbf{0} \wedge \neg x = \mathbf{1} \wedge ... \wedge \neg x = \mathbf{36} \wedge ...$, which is quite compatible with x being identical to 36. For the same reason, the truth of a statement that there exists no algorithm to solve the halting problem does not rule out its existence.

But even if all this is right, what could a physical system be like that realised the inconsistent theory of computation? After all, the Turing machine—or whatever computational device it is that is described by the inconsistent arithmetic system—is an abstract one. How this is realised physically is another matter. And if the abstract description correctly characterises the physical device, that device must be such as to render inconsistent statements such as $\exists x T_1(x, \mathbf{e}, \mathbf{i})$ and $\neg\exists x T_1(x, \mathbf{e}, \mathbf{i})$ true. How could this be?

It must be remembered that concrete devices are limited in space and time, break down and misbehave in other ways. A concrete device only ever instantiates a theoretical device imperfectly. Maybe a physical device could only approximate the abstract device up to consistency. But why should the device not behave inconsistently? Perhaps we find this difficult to imagine; but imagination is a poor index of what is possible. Notoriously, we can imagine impossible things, whilst many possible things seem hard to imagine. For example, the idea that one and the same displacement may be a spatial one, according to one observer, and a temporal one, according to another—as required by the Special Theory of Relativity—still seems hard to get one's head around in any but a mathematical fashion.

And once one moves to a paraconsistent logic, there is no *a priori* reason as to why physical reality must be consistent. Nor need this require macro-objects tables and refrigerators to behave inconsistently. The inconsistency might be purely at the unobservable level—for example, as with an electron going through distinct slits simultaneously (as would appear to be the case in the two-slit experiment in quantum mechanics). It is not even clear that an inconsistent theoretical computing device needs an inconsistent reality to encode its workings. Consider quantum computers, for example. These are devices that work with registers whose states at any time may be superpositions of classical states. Now, what is it for a machine to stop and not to stop? Simply for it to have both a terminal symbol and a non-terminal symbol in the appropriate register. (The clock, after all, does not "stop ticking". It is just that a terminal symbol, once there, stays there.) Why should these symbols not occur in a superposed state? If Schrödinger's cat can be dead and alive, so can the program be.

These last remarks are all very speculative, and in the current state of thinking in philosophy and physics, necessarily to. But they should at least serve as a warning that one cannot dismiss an inconsistent theory of computation out of hand.

6 Conclusion

In this lecture I have been looking at the implications of paraconsistency for the limitative results of metalogic and metamathematics. We have seen that once one takes the possibility of inconsistent theories seriously, much of the import of the limitative theorems of classical metamathematics is undercut. The limitative theorems of classical logic stand in even stronger forms. However, if one is prepared to countenance an inconsistent metatheory itself, even some of these may fail—and in a very spectacular way.[9]

References

[1] Meyer, R. K. (1978), 'Relevant Arithmetic', *Bulletin of the Section of Logic, Polish Academy of Sciences* 3, 133-7.

[2] Priest, G. (1987), *In Contradiction; a Study of the Transconsistent*, Dordrecht: Kluwer Academic Publishers.

[3] Priest, G. (1994), 'Is Arithmetic Consistent?', *Mind* 103, 337-349.

[4] Priest, G. (1997), Inconsistent Models of Arithmetic, I: Finite Models', *Journal of Philosophical Logic* 76, 223-235.

[5] Priest, G. (2000), Inconsistent Models of Arithmetic, II: the General Case', *Journal of Symbolic Logic* 65, 1519-1529.

[6] Priest, G. (200a), 'Paraconsistent Logic', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edition, Dordrecht: Kluwer Academic Publishers, volume to appear.

[7] Priest, G. (200b), 'Paraconsistent Set Theory and Metatheory', in A. Irvine (ed.), *Essays on Set Theory*, Oxford: Blackwells, to appear.

[8] Shapiro, S. (200a), 'Incompleteness and Inconsistency', *Mind*, to appear.

[9]Thanks go to Brad Armour-Garb and, especially, Jack Copeland for helpful discussions about the material in the last section. The possible connection between inconsistent machines and quantum computing I owe entirely to Jack.