Graham Priest

Inconsistent Arithmetics: Issues Technical and Philosophical

The study of inconsistent arithmetics is relatively young, going back about 25 years. It has, however, already occasioned a number of both interesting technical results and philosophical controversies. There are still, moreover, important technical questions to be answered, and philosophical issues to be debated. In this paper I will review the area and discuss some of these issues. In the first part of the paper I will survey the relevant technical material, ending with a number of open problems; in the second part, I will review some of the philosophical material, ending with a discussion of one central debate in the area.

1. Technical Issues

1.1. Historical Overview

The first person to construct an inconsistent arithmetic (as far as I know) was Nelson (1959), who used a realisability semantics to produce an inconsistent arithmetic, based on a paraconsistent logic of an intuitionist kind. Current developments in the subject, however, trace back, not to this, but to Meyer's paper, 'Relevant Arithmetic'. This paper, in an incomplete form, was circulated amongst relevant logicians, and was abstracted as Meyer (1976); sadly, the full version of the paper has never appeared, as far as I know. Meyer's concern was relevant Peano Arithmetic, that is, essentially, the axiomatic arithmetic in which one takes the Peano Axioms, replaces the conditionals employed with a relevant conditional, and then uses an underlying relevant logic, R in Meyer's case, to prove things about numbers. In investigating the properties of this theory, Meyer noticed that it could be given a finitary consistency proof—showing that Gödel's Second Incompleteness Theorem

V. F. Hendricks and *J. Malinowski* (eds.),
Trends in Logic: 50 Years of Studia Logica

Trends in Logic **21**: 257–283, 2003.

may not apply once one jettisons classical logic. Specifically, there are models with a two-element domain which verify all the theorems. The models were also models of the three-valued logic $RM3$, and they did a lot more than verify all the theorems of relevant Peano Arithmetic: they verified an inconsistent set of sentences.

In Meyer and Mortensen (1984) generalisations of Meyer's model were investigated.[1] Specifically, different finite sizes of the domain were employed, as were different many-valued semantics for the conditional. It thus became clear that there was a substantial family of inconsistent arithmetics. The models were constructed, in effect, by deploying a relation on the natural numbers which is a congruence with respect to successor, addition, and multiplication. Mortensen realised that similar techniques could be applied to the numbers in a non-standard model of arithmetic. In Mortensen (1986), he constructed many inconsistent arithmetics using this technique. In (1988) he investigated a number of the properties of this family of arithmetics. He also observed that the techniques in question could be applied, equally, to give inconsistent theories of other sorts of mathematical theory, such as fields and rings. In (1990) he deployed these ideas to produce inconsistent models of the differential calculus. Mortensen's work is nicely summarised in his (1995). Priest (1995), Part 4, Technical Appendix, used similar techniques to construct inconsistent set-theories with various properties. (In what follows, I will restrict myself to considering just the inconsistent natural-number arithmetics.)

Two things had become clear by this time. The first is that the inconsistent arithmetics are very powerful. Specifically, they can be made to contain all of the sentences true in the standard model of arithmetic—as expressed using just the classical propositional connectives, $\wedge$, $\vee$, $\neg$ and $\supset$, where $\alpha \supset \beta$ is defined in the usual way as $\neg\alpha \vee \beta$. The second was that, although a lot of the initial interest in these arithmetics was occasioned by an interest in a non-material conditional, and, specifically, in what could be proved using such a conditional, once one moved to a model-theoretic perspective, the non-material conditional was not playing a large role: all of the truths of the standard model came for free anyway. This meant that one could simply forget about the non-material conditional, and investigate the structure of the theories, as expressed in the classical vocabulary (though the underlying logic could not, of course, be classical, since the interpretations model inconsistent sets of sentences). This, in turn, allowed inconsistent models to

[1] In Mortensen and Meyer (1985) there is also an application of the inconsistent models to arithmetic based on a non-distributive quantum logic.

be constructed by a simple yet powerful model-theoretic construction now called the Collapsing Lemma. A form of this had already been established by Dunn in (1979); a version which allowed it to be immediately applicable to the construction of inconsistent models was given in Priest (1991).

All the tools were now at hand for circumscribing an important class of inconsistent arithmetics, and investigating their structure. The class was simple enough to be natural, and complex enough to be mathematically interesting. The analysis of the finite case was given in Priest (1997), and of the general case in Priest (2000). I will explain the details in what follows.

1.2. LP and FDE

The underlying logic for the arithmetics is First Degree Entailment (FDE) or LP. It does not make a difference which, as we will see in a moment. Let me start by specifying the semantics of the logic.[2]

The vocabulary of the languages in question is that of first-order logic with identity. (We take $\supset$ to be defined in the usual way.) For simplicity, I suppose that there are no free variables. An FDE interpretation, $\mathcal{I}$, for the language is a *pair*, $\langle D, d\rangle$, where D is the (non-empty) domain of quantification, and d is a function that maps every constant to an object in the domain, every n-place function symbol to an n-place function on the domain, and every n-place predicate, P, to a pair, $\langle E_P, A_P\rangle$, each member of which is a subset of the set of n-tuples of D, D^n. E_P is the *extension* of P; A_P is the *anti-extension.* The extension of the identity predicate, $E_=$, is, as usual, $\{\langle d, d\rangle : d \in D\}$. For an LP interpretation, we require, in addition, that for every n-place predicate, P, $E_P \cup A_P = D^n$. (This is the only difference between FDE and LP.)

Every term, t, is assigned a denotation, $d(t)$, in the usual recursive fashion. Every sentence, α, is assigned a truth value, $\nu(\alpha)$, which is a subset of $\{1, 0\}$, non-empty in the case of LP. For atomic sentences, the truth/falsity conditions are:

$1 \in \nu(Pt_1...t_n)$ iff $\langle d(t_1), ..., d(t_n)\rangle \in E_P$
$0 \in \nu(Pt_1...t_n)$ iff $\langle d(t_1), ..., d(t_n)\rangle \in A_P$

Truth/falsity conditions for the connectives are:

$1 \in \nu(\neg\alpha)$ iff $0 \in \nu(\alpha)$
$0 \in \nu(\neg\alpha)$ iff $1 \in \nu(\alpha)$

$1 \in \nu(\alpha \wedge \beta)$ iff $1 \in \nu(\alpha)$ and $1 \in \nu(\beta)$

[2]For further details, see Priest (2001), ch. 8.

$0 \in \nu(\alpha \wedge \beta)$ iff $0 \in \nu(\alpha)$ or $0 \in \nu(\beta)$

$1 \in \nu(\alpha \vee \beta)$ iff $1 \in \nu(\alpha)$ or $1 \in \nu(\beta)$
$0 \in \nu(\alpha \wedge \beta)$ iff $0 \in \nu(\alpha)$ and $0 \in \nu(\beta)$

For the quantifiers, we assume that the language has been augmented, if necessary, by names, in such a way that every member, d, of D has a name, $\mathbf{d}$.

$1 \in \nu(\forall x\alpha(x))$ iff for every $d \in D$, $1 \in \nu(\alpha(\mathbf{d}))$
$0 \in \nu(\forall x\alpha(x))$ iff for some $d \in D$, $0 \in \nu(\alpha(\mathbf{d}))$

$1 \in \nu(\exists x\alpha(x))$ iff for some $d \in D$, $1 \in \nu(\alpha(\mathbf{d}))$
$0 \in \nu(\exists x\alpha(x))$ iff for every $d \in D$, $0 \in \nu(\alpha(\mathbf{d}))$

An interpretation is a *model* of α iff, in that interpretation, $1 \in \nu(\alpha)$; it is a model of a set of sentences, Σ, iff it is a model of every member of Σ; and an inference is *valid* ($\vDash$) iff every model of the premises is a model of the conclusion.

It should be noted that if, for every predicate, P, E_P and A_P are exclusive and exhaustive, then we have, in effect, an interpretation of classical first order logic. All classical interpretations are therefore FDE (and LP) interpretations.

1.3. The Collapsing Lemma

We can now state the Collapsing Lemma. Let $\mathcal{I} = \langle D, d \rangle$ be any interpretation. Let $\sim$ be an equivalence relation on D, which is also a congruence relation on the denotations of the function symbols in the language (i.e., if g is such a denotation, and $d_i \sim e_i$ for all $1 \leq i \leq n$, then $g(d_1, ..., d_n) \sim g(e_1, ..., e_n)$). If $d \in D$ let $[d]$ be the equivalence class of d under $\sim$. Define an interpretation, $\mathcal{I}_\sim = \langle D_\sim, d_\sim \rangle$, to be called the *collapsed interpretation*. $D_\sim = \{[d]; d \in D\}$; if c is a constant, $d_\sim(c) = [d(c)]$; if f is an n-place function symbol:

$$d_\sim(f)([d_1], ..., [d_n]) = [d(f)(d_1, ..., d_n)]$$

(this is well defined, since $\sim$ is a congruence relation); and if P is an n-place predicate, its extension and anti-extension in $\mathcal{I}_\sim$, $E_P^\sim$ and $A_P^\sim$, are defined by:

$\langle [d_1], ..., [d_n] \rangle \in E_P^\sim$ iff for all $1 \leq i \leq n$, $\exists e_i \sim d_i$, $\langle e_1, ..., e_n \rangle \in E_P$
$\langle [d_1], ..., [d_n] \rangle \in A_P^\sim$ iff for all $1 \leq i \leq n$, $\exists e_i \sim d_i$, $\langle e_1, ..., e_n \rangle \in A_P$

where E_P and A_P are the extension and anti-extension of P in $\mathcal{I}$. It is easy to check that $E_=^\sim$ is $\{\langle [d], [d] \rangle ; d \in D\}$, as required.

The collapsed interpretation, in effect, identifies all members of an equivalence class to produce a composite individual which has the properties of all of its members. It may, of course, be inconsistent, even if its members are not.

A swift induction confirms that for any term, t, $d_\sim(t) = [d(t)]$. Hence:

$$\begin{aligned} 1 \in \nu(Pt_1...t_n) &\Rightarrow \langle d(t_1), ..., d(t_n)\rangle \in E_P \\ &\Rightarrow \langle [d(t_1)], ..., [d(t_n)]\rangle \in E_P^\sim \\ &\Rightarrow \langle d_\sim(t_1), ..., d_\sim(t_n)\rangle \in E_P^\sim \\ &\Rightarrow 1 \in \nu_\sim(Pt_1...t_n) \end{aligned}$$

Similarly for 0 and anti-extensions. A routine argument by induction then establishes that this condition obtains for all formulas; i.e., for all α, $\nu(\alpha) \subseteq \nu_\sim(\alpha)$. This is the *Collapsing Lemma*.[3]

The Collapsing Lemma assures us that if an interpretation is a model of some set of sentences, then any interpretation obtained by collapsing it will also be a model. This gives us an important way of constructing inconsistent models of a theory. Start with any model of the theory, possibly a classical model, and collapse. Any collapsed model will be a model of the original theory, though, in general, it will be inconsistent. I will illustrate by constructing some inconsistent models of arithmetic.

1.4. Collapsed Models of Arithmetic

Let us start with a definition. Let L be the standard language of first-order arithmetic: one constant, $\mathbf{0}$, function symbols for successor, addition, and multiplication ($'$, $+$, and $\times$, respectively), and one predicate symbol, $=$. Let $\mathcal{N}$ be the standard (classical) interpretation of this language; and if $\mathcal{M}$ is an interpretation, let $Th(\mathcal{M})$ be the set of sentences true in $\mathcal{M}$. A *model of arithmetic* is any FDE or LP interpretation of L which is a model of $Th(\mathcal{N})$. Note that since $Th(\mathcal{N})$ is complete (i.e., for all α, either $\alpha \in Th(\mathcal{N})$ or $\neg\alpha \in Th(\mathcal{N})$), an FDE model is also an LP model. Note, also, that, as well as $\mathcal{N}$, any classical non-standard model of arithmetic is a model of arithmetic in the sense I will use the word here. But there are many more. In particular, as we will see, there are models of arithmetic, $\mathcal{M}$, such that $Th(\mathcal{M})$ is inconsistent. I will call such models, naturally enough, inconsistent models of arithmetic.

Now, let $\mathcal{M} = \langle M, d\rangle$ be any classical model of $Th(\mathcal{N})$. Let $\sim$ be an equivalence relation on M which is also a congruence relation with respect to the interpretations of the function symbols. Then we may construct the

[3] For details of the proof, see Priest (1991).

collapsed interpretation, $\mathcal{M}_\sim$. By the Collapsing Lemma, $\mathcal{M}_\sim$ is a model of arithmetic. Provided that $\sim$ is not the trivial equivalence relation, that relates each thing to only itself, then $\mathcal{M}_\sim$ will model inconsistencies. For suppose that $\sim$ relates the distinct members of M, n and m, then in $\mathcal{M}_\sim$, $[n] = [m]$ and so $\langle [n], [m] \rangle$ is in the extension of $=$. But since $n \neq m$ in $\mathcal{M}, \langle [n], [m] \rangle$ is in the anti-extension too. Thus, $\exists x(x = x \wedge x \neq x)$ holds in $\mathcal{M}_\sim$. Let me give a couple of simple examples of this.

Example one. Let $\mathcal{M}$ be the standard model of arithmetic. n, $p \in M$ and $p > 0$. Define a relation, $\sim$, on M, thus:

$$x \sim y \text{ iff } (x,\, y < n \text{ and } x = y) \text{ or } (x,\, y \geq n \text{ and } x = y \pmod{p})$$

It is easy to check that $\sim$ is a congruence relation on M. Let $\mathcal{M}_n^p$ be the model obtained by collapsing with respect to this. The Collapsing Lemma assures us that it is a model of arithmetic. It is finite; it has an initial tail of length n that behaves consistently. The other numbers form a cycle of period p. The successor graph can be depicted as follows:

$$\begin{array}{ccccccccc} 0 & \rightarrow & 1 & \rightarrow & ... & \rightarrow & n & \rightarrow & n+1 \\ & & & & & & \uparrow & & \downarrow \\ & & & & & & n+p-1 & \leftarrow & \cdots \end{array}$$

Example two. Let $\mathcal{M}$ be any non-standard classical model of arithmetic. Define the relation $\sim$ as follows:

$$x \sim y \text{ iff } (x,\, y \text{ are standard numbers and } x = y) \text{ or } (x,\, y \text{ are non-standard})$$

Again, it is easy to check that $\sim$ is an equivalence relation which is also a congruence on the arithmetic operators. The model obtained by collapsing with respect to this equivalence relation contains the standard interpretation, plus an inconsistent "point at infinity". The successor graph can be depicted as follows:

$$\begin{array}{ccccccc} 0 & \rightarrow & 1 & \rightarrow & ... & \Omega \\ & & & & & \circlearrowright \end{array}$$

1.5. Inconsistent Models of Arithmetic

We now turn to the question of the general structure of models of arithmetic.[4] Let $\mathcal{M} = \langle M, d \rangle$ be any such model. I will refer to the denotations of $'$, $+$,

[4]The material in this and the next section is reproduced with minor revisions from sections 9.3, 9.4 of Priest (2002). I am grateful for permission to reuse the material. The contents of these sections are covered in more detail in Priest (2000).

and $\times$ as the arithmetic operations of $\mathcal{M}$; and since no confusion is likely, use the same signs for them. I will call the denotations of the numerals *regular* numbers.

Let $x \leq y$ be defined, in the usual way, as $\exists z\ x + z = y$. It is easy to check that $\leq$ is transitive. For if $i \leq j \leq k$, then for some x, y, $i + x = j$ and $j + y = k$. Hence $(i + x) + y = k$. But $(i + x) + y = i + (x + y)$ (since we are dealing with a model of arithmetic). The result follows.

If $i \in M$, let $N(i)$ (the *nucleus* of i) be $\{x \in M; i \leq x \leq i\}$. In a classical model, $N(i) = \{i\}$, but this need not be the case in an inconsistent model. For example, in any $\mathcal{M}_n^p$ the members of the cycle constitute a nucleus. If $j \in N(i)$ then $N(i) = N(j)$. For if $x \in N(j)$ then $i \leq j \leq x \leq j \leq i$, so $x \in N(i)$, and similarly in the other direction. Thus, every member of a nucleus defines the same nucleus.

Now, if N_1 and N_2 are nuclei, define $N_1 \preceq N_2$ to mean that for some (or all, it makes no difference) $i \in N_1$ and $j \in N_2$, $i \leq j$. It is not difficult to check that $\preceq$ is a partial ordering. Moreover, since for any i and j, $i \leq j$ or $j \leq i$, it is a linear ordering. The least member of the ordering is $N(0)$. If $N(1)$ is distinct from this, it is the next (since for any x, $x \leq 0 \vee x \geq 1$), and so on for all regular numbers.

Say that $i \in M$ has *period* $p \in M$ iff $i + p = i$. In a classical model every number has period 0 and only 0. But again, this need not be the case in an inconsistent model, as the $\mathcal{M}_n^p$ demonstrate. If $i \leq j$ and i has period p, so does j. For $j = i + x$, so $p + j = p + i + x = i + x = j$. In particular, if p is a period of some member of a nucleus, it is a period of every member. We may thus say that p is a period of the nucleus itself. It also follows that if $N_1 \preceq N_2$ and p is a period of N_1 it is a period of N_2.

If a nucleus has a regular non-zero period, m, then it must have a minimum (in the usual sense) non-zero period, since the sequence $0, 1, 2, ..., m$ is finite. If $N_1 \preceq N_2$ and N_1 has minimum regular non-zero period, p, then p is a period of N_2. Moreover, the minimum non-zero period of N_2, q, must be a divisor (in the usual sense) of p. For suppose that $q < p$, and that q is not a divisor of p. For some $0 < k < q$, p is some finite multiple of q plus k. So if $x \in N_2$, $x = x + q = x + p + ... + p + k$. Hence $x = x + k$, i.e., k is a period of N_2, which is impossible.

If a nucleus has period $p \geq 1$, I will call it *proper*. Every proper nucleus is closed under successors. For suppose that $j \in N$ with period p. Then $j \leq j' \leq j + p = j$. Hence, $j' \in N$. In an inconsistent model, a number may have more than one predecessor, i.e., there may be more than one x

such that $x' = j$.[5] (Although $(x' = y') \supset x = y$ holds in the model, we cannot necessarily detach to obtain $x = y$.) But if j is in a proper nucleus, N, it has a unique predecessor in N. For let the period of N be q'. Then $(j + q)' = j + q' = j$. Hence, $j + q$ is a predecessor of j; and $j \leq j + q \leq j + q' = j$. Hence, $j + q \in N$. Next, suppose that x and y are in the nucleus, and that $x' = y' = j$. We have that $x \leq y \vee y \leq x$. Suppose, without loss of generality, the first disjunct. Then for some z, $x + z = y$; so $j + z = j$, and z is a period of the nucleus. But then $x = x + z = y$. I will write the unique predecessor of j in the nucleus as $'j$.

Now let N be any proper nucleus, and $i \in N$. Consider the sequence $...,'' i, ' i, i, i', i''....$ Call this the *chromosome* of i. Note that if i, $j \in N$, the chromosomes of i and j are identical or disjoint. For if they have a common member, z, then all the finite successors of z are identical, as are all its finite predecessors (in N). Thus they are identical. Now consider the chromosome of i, and suppose that two members are identical. There must be members where the successor distance between them is a minimum. Let these be j and $j'^{\cdots'}$ where there are n primes. Then $j = j + n$, and n is a period of the nucleus—in fact, its minimum non-zero period—and the chromosome of every member of the nucleus is a successor cycle of period n.

Hence, any proper nucleus is a collection of chromosomes, all of which are either successor cycles of the same finite period, or are sequences isomorphic to the integers (positive and negative). Both sorts are possible in an inconsistent model. Just consider the collapse of a non-standard model, of the kind given in 1.4, by an equivalence relation which leaves all the standard numbers alone and identifies all the others modulo p. If p is standard, the non-standard numbers collapse into a successor cycle; if it is non-standard, the nucleus generated is of the other kind.

To summarise so far, the general structure of a model is a liner sequence of nuclei. There are three segments (any of which may be empty). The first contains only improper nuclei. The second contains proper nuclei with linear chromosomes. The final segment contains proper nuclei with cyclical chromosomes of finite period. A period of any nucleus is a period of any subsequent nucleus; and in particular, if a nucleus in the third segment has minimum non-zero period, p, the minimum non-zero period of any subsequent nucleus is a divisor of p. Thus, we might depict the general structure

[5] In fact, it is not difficult to show that there is at most one number with multiple predecessors; and this can have only two.

of a model as follows (where $m+1$ is a multiple of $n+1$):

$$0,1,\ldots \quad \boxed{\begin{array}{c} \ldots a \to a' \ldots \\ \ldots b \to b' \ldots \\ \vdots \end{array}} \quad \ldots \quad \boxed{\begin{array}{ccc} d_0 \ldots d_i & e_0 \ldots e_i & \\ \uparrow \quad \downarrow & \uparrow \quad \downarrow & \ldots \\ d_m \ldots d'_i & e_m \ldots e'_i & \end{array}} \quad \ldots \quad \boxed{\begin{array}{ccc} f_0 \ldots f_i & g_0 \ldots g_i & \\ \uparrow \quad \downarrow & \uparrow \quad \downarrow & \ldots \\ f_n \ldots f'_i & g_n \ldots g'_i & \end{array}} \quad \ldots$$

An obvious question at this point is what orderings the proper nuclei may have. For a start, they can have the order-type of any ordinal. To prove this, one establishes by transfinite induction that for any ordinal, α, there is a classical model of arithmetic in which the non-standard numbers can be partitioned into a collection of disjoint blocks with order-type α, closed under arithmetic operations. One then collapses this interpretation in such a way that each block collapses into a nucleus.

The proper nuclei need not be discretely ordered. They can also have the order-type of the rationals. To prove this, one considers a classical non-standard model of arithmetic, where the order-type of the non-standard numbers is that of the rationals. It is then possible to show that the numbers can be partitioned into a collection of disjoint blocks, closed under arithmetic operations, which themselves have the order-type of the rationals. One can then collapse this model in such a way that each of the blocks collapses into a proper nucleus, giving the result. This proof can be extended to show that any order-type which can be embedded in the rationals in a certain way can also be the order-type of the proper nuclei. This includes ω^* (the reverse of ω) and $\omega^* + \omega$, but *not* $\omega + \omega^*$.

1.6. Finite Models of Arithmetic

First-order arithmetic has many classical non-standard models, but none of them is finite. One of the intriguing features of inconsistent models of arithmetic is that they can be just that, e.g., the $\mathcal{M}_n^p$. For finite models, a complete characterisation is known.

Placing the constraint of finitude on the results of 1.5, we can infer as follows. The sequence of improper nuclei is either empty or is composed of the singletons of $0, 1, \ldots, n$, for some finite n. There must be a finite collection of proper nuclei, $N_1 \preceq \ldots \preceq N_m$; each N_i must comprise a finite collection of successor cycles of some minimum non-zero finite period, p_i. And if $1 \leq i \leq j \leq m$, p_j must be a divisor of p_i.[6]

Moreover, there are models of any structure of this form. To show this, take any non-standard classical model of arithmetic. This can be partitioned

[6] It is also possible to show that each nucleus is closed under addition and multiplication.

into the finite collection of blocks:

$$C_0, C_{1_0}, ..., C_{1_{k(1)}}, ...C_{i_0}, ..., C_{i_{k(i)}}, ..., C_{m_0}, ..., C_{m_{k(m)}}$$

where C_0 is either empty or is of the form $\{0, ..., n\}$, and each subsequent block is closed under arithmetic operations.[7] We now define a relation, $x \sim y$, as follows:

$(x, y \in C_0$ and $x = y)$ or
for some $1 \leq i \leq m$:
(for some $0 < j < k(i)$, $x, y \in C_{i_j}$, and $x = y \bmod p_i$) or
$(x, y \in C_{i_0} \cup C_{i_{k(i)}}$ and $x = y \bmod p_i)$

One can check that $\sim$ is an equivalence relation, and also that it is a congruence relation on the arithmetic operations. Hence we can construct the collapsed model. $\sim$ leaves all members of C_0 alone. For every i it collapses every C_{i_j} into a successor cycle of period p_i, and it identifies the blocks C_{i_0} and $C_{i_{k(i)}}$. Thus, the sequence $C_{i_0}, ..., C_{i_{k(i)}}$ collapses into a nucleus of period p_i with $k(i)$ chromosomes. The collapsed model therefore has exactly the required structure.

Finally in this section, note the following. If $\mathcal{M}$ is finite, $Th(\mathcal{M})$ is decidable. The truth value(s) in $\mathcal{M}$ of an atomic sentence can be computed, since the denotations of the functions and predicates, being on a finite domain, are computable. The truth values of propositional compounds are computed by LP (FDE) truth tables, and, since the domain is finite, quantified sentences are equivalent to finite conjunctions/disjunctions. (Thus, $\exists x\alpha(x)$ has the same truth value(s) as $\alpha(\mathbf{d}_0) \vee ... \vee \alpha(\mathbf{d}_n)$, where $D = \{d_0, ..., d_n\}$.)

1.7. Open Problems

There are many interesting questions about inconsistent models, even the finite ones, the answers to which are not known. I finish the technical part of this paper by listing some.

- Characterise the orderings of the proper cycles that may be realised in an inconsistent model.
- Can a nucleus have an infinitely descending sequence of periods?
- Given a model with a particular structure of cycles (nuclei and chromosomes), how many models of that structure are there? (The behaviour

[7] The existence of such a partition follows from a standard result in the study of classical models of arithmetic. See Kaye (1991), sect. 6.1.

of the successor function in a model does not determine the behavior of addition and multiplication, except in the tail.)

- Must nuclei always be closed under addition and multiplication?

- The set of sentences true in any finite model is decidable, and *a fortiori* axiomatisable. Are there any infinite models such that the set of sentences that hold in them is axiomatic?

Perhaps the most fundamental open question is as follows. Not all inconsistent models of arithmetic are collapses of classical models. Let $\mathcal{M}$ be any model of arithmetic; if $\mathcal{M}'$ is obtained from $\mathcal{M}$ by adding extra pairs to the anti-extension of $=$, call $\mathcal{M}'$ an *extension* in $\mathcal{M}$. If $\mathcal{M}'$ is an extension of $\mathcal{M}$, $\mathcal{M}'$ is itself a model of arithmetic (as may be shown by a simple inductive argument). Now, consider the extension of the standard model obtained by adding $\langle 0, 0\rangle$ to the anti-extension of $=$. This is not a collapsed model, since, if it were, 0, being inconsistent, would have to have been identified with some $x > 0$. But then 1 would have been identified with $x' > 1$. Hence, $\mathbf{0}' \neq \mathbf{0}'$ would also be true in the model, which it is not. Maybe, however, the following conjecture is true:

- Each inconsistent model is the extension of a collapsed classical model.

If this conjecture is correct, collapsed models can be investigated via an analysis of the classical models of arithmetic and their congruence relations.

2. Philosophical Issues

2.1. Historical Overview

We now turn to some issues in the philosophy of arithmetic which are posed by the existence of inconsistent models. I start, as with the technical section, by giving a brief historical overview.

The first paper to argue for the inconsistency of arithmetic predates the technical investigations, and was Priest (1979). The argument appeals to Gödel's first Incompleteness Theorem. This was criticised by Chihara (1984). An extended defence of the argument appeared in Priest (1987), ch. 3.

The first person to deploy the technical material on inconsistent models in a philosophical context was van Bendegem (1993), (1994). He was particularly concerned with the finite models, and developed an argument for finitism on the basis of them. Priest (1994a) took up the idea, but used the finite models in defence of inconsistent arithmetic, rather than finitism. Where van Bendegem saw a greatest number, Priest saw a least inconsistent number. The idea of a least inconsistent number was discussed further in Priest (1994b). Priest (1994a) invokes various results about the finite models and metatheoretic notions, particularly provability. The relationship between finite models and provability is discussed further in Mortensen (1995). Priest and van Bendegem's deployment of inconsistent models is criticised by Batens (200a). Denyer (1995) is also a critique of Priest, to which Priest (1996a) is a reply.

Priest (1996b) invokes the inconsistent models of arithmetic in a quite different way: to argue the case for the possibility of arithmetic revision.[8] Most recently, Priest's (1987) argument for the inconsistency of arithmetic has been criticised by Shapiro (2002), though the technical material on inconsistent models is not deployed.

In the second half of this paper I will review and discuss some of these developments. At the end, I will address some of Shapiro's criticisms specifically in the light of the technical material on inconsistent arithmetics.

Let me start by setting the scene. If $\mathcal{M}$ is any model of arithmetic $Th(\mathcal{M})$ is a theory, that is, a set of arithmetic sentences closed under LP (and FDE) consequence, and contains $Th(\mathcal{N})$. If $\mathcal{M}$ is an inconsistent model of arithmetic, $Th(\mathcal{M})$ is also inconsistent.[9] I will call any such theory an *inconsistent arithmetic*. Now, when we count and perform arithmetic operations, which theory of arithmetic is right?

2.2. Arithmetic Revision

To answer this question we need to distinguish between pure arithmetic and applied arithmetic. A pure arithmetic is the set of truths about numbers themselves. An applied arithmetic is a pure arithmetic employed for the purposes of counting something or other. Now, if one askes what the correct pure arithmetic is, a natural answer is that the correct theory of arithmetic is $Th(\mathcal{N})$, the set of sentences true in the "standard model". Let us, for the

---

[8]This paper was due to appear in the proceedings of the conference at which it was given. These never, unfortunately, eventuated. It is worth noting that the idea that one might use a non-standard arithmetic to count appears as early as Gasking (1940).

[9]There are, in fact, LP theories that contain all of $Th(\mathcal{N})$, but that are not the theory of some collapsed model. This, for example, $\bigcap_{n\in\omega}\mathcal{M}^p_n$, being an intersection of theories, is a theory. But it contains the sentence $\exists x(x \neq x)$, whilst it contains nothing of the form $\mathbf{n} \neq \mathbf{n}$.

moment, agree that this is true.

It remains the case that, at least for some purposes, we might wish to use a different arithmetic to count some things. Compare the situation with that in geometry. Until the 19th century, the correct geometry for application to space was taken to be Euclidean geometry. But in the 20th century, this position was revised. The correct geometry is not Euclidean, but some non-Euclidean geometry. Could it be that we might want to revise our arithmetic in the same way?

It seems to me that we might. As fallibilists have argued, any theory that we employ may be revised under the pressure of recalcitrant evidence. Mathematical theories are no exception—as the history of geometry demonstrates. Whilst I have no situation to offer where the revision of arithmetic is currently plausible, it is easy enough to imagine the possibility of such things. Here, for example, is one where we might be inclined to revise our arithmetic in favour of $Th(\mathcal{M}^1_n)$ (where, for all $m \geq n$, $m = n$.)

Let us suppose that we come to predict a collision between an enormous star and a huge planet.[10] Using a standard technique, we compute their masses as x_1 and y_1, respectively. Since masses of this kind are, to within experimental error, the sum of the masses of the baryons (protons and neutrons) in them, it will be convenient to take a unit of measurement according to which a baryon has mass 1. In effect, therefore, these figures measure the numbers of baryons in the masses. After the collision, we measure the mass of the resulting (fused) body, and obtain the figure n, where n is much less than $x_1 + y_1$. Naturally, our results are subject to experimental error. But the difference is so large that it cannot possibly be explained by this. We check our instruments, suspecting a fault, but cannot find one; we check our computations for an error, but cannot find one. We have a puzzle. Some days later, we have the chance to record another collision. We record the masses before the collision. This time they are x_2 and y_2. Again, after the collision, the mass appears to be n (the same as before), less than $x_2 + y_2$. The first result was no aberration. We have an anomaly.

We investigate various ways of solving the anomaly. We might revise the theories on which our measuring devices depend, but there is no obvious way of doing this. We could say that some baryons disappeared in the collision; alternatively, we could suppose that under certain conditions the mass of a baryon decreases. But either of these options seems to amount to a rejection of the law of conservation of mass(-energy), which would seem to be a rather unattractive course of action.

[10]The following example comes, with minor revisions, from Priest (1996b).

Then someone, call them Einquine, fixes on the fact that the resultant masses of the two collisions were the same in both cases, n. This is odd. If mass has gone missing, why should this produce the same result in both cases? An idea occurs to Einquine. Maybe our arithmetic for counting baryons is wrong.[11] Maybe the appropriate arithmetic is $Th(\mathcal{M}^1_n)$. For in this arithmetic $x_1 + y_1 = x_2 + y_2 = n$, and our observations are explained without having to assume that the mass of baryons has changed, or that any are lost in the collisions! Einquine hypothesizes that n is a fundamental constant of the universe, just like the speed of light, or Planck's constant.[12]

While she is thus hypothesising, reports of the collisions start to come in from other parts of the galaxy. (The human race had colonised other planets some centuries before.) These reports all give the masses of the two new objects as the same, but all are different from each other. Some even measure them as greater that the sum of their parts. Einquine is about to give up her hypothesis, when she realises that this is quite compatible with it. Even if the observer measures the mass as m, provided only that $m \geq n$ then $m = n$, and their results are the same!

But this does leave a problem. Why do observers consistently record numerical results that differ from each other? Analysing the data, Einquine sees that values of n (hers included), are related to the distance of the observer from the collision, d, by the (classical) equation $n = n_0 + kd$ (where n_0 and k are constants). In virtue of this, she revises her estimate of the fundamental constant to n_0, and hypothesizes that the effect of an inconsistent mass of baryons on a measuring device is a function of its distance from the mass. Further observational reports bear this hypothesis out; and Einquine starts to consider the mechanism involved in the distance-effect.

We could continue the story indefinitely, but it has gone far enough. For familiar reasons, there are likely to be theories other than Einquine's that could be offered to explain the data. Some of them might preserve orthodox arithmetic by jettisoning conservation laws, or by keeping these but varying some physical auxiliary hypotheses. Others might modify arithmetic in some other, but consistent, way. And each of these theories might become more or less plausible in the light of further experimentation, etc. But the point is made: it is quite possible that we might vary our arithmetic for empirical

[11] We already know that different sorts of fundamental particles satisfy different sorts of statistics.

[12] The revision of arithmetic envisaged here is a local one, in that it is only the counting of *baryons* that is changed. It would be interesting to speculate on what might happen which could motivate a global change, i.e., a move to a situation where everything is counted in the new way.

reasons. There can be alternative applied arithmetics, just as there are alternative applied geometries.

2.3. Consistent vs. Inconsistent Arithmetics

But now let us consider a stronger possibility—not just that we might want to apply an inconsistent arithmetic for some purpose, but that the correct pure arithmetic is one of the inconsistent ones. Which one? That is obviously an important question; but for the present, it does not need to be addressed.[13] The following considerations do not depend on which inconsistent arithmetic is at issue—or if they do, I will make this explicit.

The orthodox view is certainly that $Th(\mathcal{N})$ is the true arithmetic, not $Th(\mathcal{M})$, where $\mathcal{M}$ is some inconsistent model of arithmetic. Of course $Th(\mathcal{N})$ is true of $\mathcal{N}$, and $Th(\mathcal{M})$ is true of $\mathcal{M}$. That is not contentious. The question, then, is whether it is $\mathcal{N}$ or $\mathcal{M}$ that is the correct interpretation of the language. It might seem as though it is easy to resolve this issue, but it is not. A dispute between the proponent, A, of "standard arithmetic" and the proponent, B, of an inconsistent arithmetic is of a somewhat unusual kind. Anything (at least, anything arithmetic) that A endorses, B will endorse too. Thus, for example, A will insist that there is no greatest number ($\forall x \exists y\, y > x$); B will concur. The locus of disagreement will be in the fact that B will assert things that A will not wish to assent to. Why suppose A right and B wrong? A may point out that B's view of arithmetic is inconsistent; but unless they have some independent reason to suppose that inconsistency—or at least inconsistency in arithmetic—is a bad thing, this simply begs the question. A may, of course, attempt to mount a defence of consistency in general. I do not wish to enter into that debate here. Let me just say, for the record, that I am not aware of any very persuasive—and, in particular, non-question-begging—arguments for that conclusion.[14]

Are there any reasons, however, that push us towards endorsing an inconsistent arithmetic? One reason is that inconsistent arithmetics avoid some of the limitative results of the classical metatheory of arithmetic, and the unhappinesses associated with these.[15] Inconsistent arithmetics can do lots of things that consistent arithmetics cannot do. Thus, for example, as I have already noted, some inconsistent arithmetics are decidable. If one

[13] The question is discussed in the context of the finite models in Priest (1994b).

[14] See, e.g., Priest (1998).

[15] This matter is discussed further in Priest (1994a). I am not now happy with a number of the arguments used in that paper. Some reasons why are explained in Priest (1996a).

of these is the correct arithmetic then there is an algorithm for solving any arithmetical problem, which would certainly be very nice.

Another thing that inconsistent arithmetics can do is contain their own truth predicate; hence Tarski's theorem is avoided. Tarski's theorem shows that any theory that contains its own truth predicate is inconsistent; but this is obviously no problem in an inconsistent arithmetic! The language of arithmetic that we have been dealing with so far contains no truth predicate. However, it is well known[16] that any arithmetic based on LP can be extended conservatively with a truth predicate, T, satisfying the two way rule:

$$\frac{T\langle\alpha\rangle}{\overline{\alpha}}$$

where $\langle\alpha\rangle$ is the numeral of the gödel number of α.[17]

Of course, since the extension of the language with a truth predicate is conservative, if we start with a consistent arithmetic, the purely arithmetic fragment of the theory with the truth predicate will also be consistent. So the inconsistency generated by the truth predicate gives no reason, as such, to suppose that the purely arithmetic fragment is inconsistent. But if one can have a truth predicate, excluding it from "pure arithmetic" is somewhat arbitrary. Truth has just as good a claim to be considered a logical predicate as the identity predicate. It should, therefore, be a part of all "pure theories".

2.4. Gödel's Theorems

Another thing that consistent arithmetic cannot do is provide a complete axiomatic theory. Inconsistent arithmetics can do this. As I have already noted, there are decidable complete inconsistent arithmetics; *a fortiori* they are axiomatic (and so, to point out the obvious, they can be specified by an axiom system in the usual way, quite independently of any consideration of collapsed models). In virtue of the methodological importance of axiomatisability in mathematics, this is a significant plus.

---

[16]See, e.g., Priest (2002), 8.1.

[17]It is worth noting also that any finite LP model of arithmetic will model all instances of the Induction Schema, however the language of arithmetic is extended. The Schema is of the form: $(\alpha(\mathbf{0}) \wedge \forall x(\alpha(x) \supset \alpha(x'))) \supset \forall x\alpha(x)$. With a little massaging, this can be seen to be equivalent to: $\neg\alpha(\mathbf{0}) \vee \exists x(\alpha(x) \wedge \neg\alpha(x')) \vee \forall x\alpha(x)$. Now, if the last disjunct is true, we are home. If not, there is some n such that $\alpha(\mathbf{n})$ fails, and since there is only a finite number of numbers in the domain, a least such n. Since $\alpha(\mathbf{n})$ fails, $\neg\alpha(\mathbf{n})$ holds. Thus, if $n = 0$ the first disjunct holds. If not, $n = m'$, and $\alpha(\mathbf{m})$, so the middle disjunct holds.

Naturally, it is worth asking what happens to the "Gödel undecidable sentence" in these arithmetics. Take any inconsistent arithmetic, Θ, which is a collapsed model of a classical model of arithmetic, $\mathcal{M}$. Since it is axiomatisable, its membership can be represented in the standard model—and also $\mathcal{M}$, since this is elementarily equivalent to it—by a formula of one free variable, $B(x)$. ($B(x)$ is of the form $\exists y\, Prov(y, x)$, where $Prov(y, x)$ represents the proof relation in $\mathcal{M}$.) That is, for any sentence, α:[18]

If $\alpha \in \Theta$ then $B\langle\alpha\rangle \in Th(\mathcal{M})$.

If $\alpha \notin \Theta$ then $\neg B\langle\alpha\rangle \in Th(\mathcal{M})$.

By the Collapsing Lemma, it follows that:

(1) If $\alpha \in \Theta$ then $B\langle\alpha\rangle \in \Theta$.

(2) If $\alpha \notin \Theta$ then $\neg B\langle\alpha\rangle \in \Theta$.

The undecidable sentence is a sentence, γ, of the form $\neg B\langle\gamma\rangle$. It is not difficult to see that both γ and $\neg\gamma$ are provable in Θ. For either $\gamma \in \Theta$ or $\gamma \notin \Theta$. In the first case, $B\langle\gamma\rangle = \neg\gamma \in \Theta$, by (1). In the second case, $\neg\gamma \in \Theta$, since Θ is complete, but $\neg B\langle\gamma\rangle = \gamma \in \Theta$, by (2). Either way, $\gamma \wedge \neg\gamma \in \Theta$. Note that, unlike the case of the contradiction connected with Tarski's Theorem, γ is a purely arithmetic sentence; that is, its vocabulary is just that of the pure language of arithmetic.

Given the inconsistency of the arithmetic in question, a consistency proof for it, and *a fortiori* a consistency proof within Θ, is not to be expected. Classically, of course, consistency and non-triviality go together; but in a paraconsistent context, this is not the case. In particular, though Θ is inconsistent, it is not trivial (unless it is produced by collapsing under the degenerate equivalence relation that relates everything to everything). And the non-triviality of Θ can be proved within Θ. In this sense, Gödel's second Incompleteness Theorem fails for inconsistent arithmetics. For take any unprovable sentence, α. Then since $\alpha \notin \Theta$, $\neg B\langle\alpha\rangle \in \Theta$, by (2). (Beware, however. This does not rule out $B\langle\alpha\rangle$ from being in Θ, too! We will return to this matter later.)

Finally, closely connected with Gödel's second Incompleteness Theorem is Löb's Theorem, to the effect that in classical arithmetics if $B\langle\alpha\rangle \supset \alpha$ is provable, so is α. It follows that not all instances of $B\langle\alpha\rangle \supset \alpha$ are provable. But this seems odd. All such sentences are clearly true; how is that truths

[18] I write $B\langle\alpha\rangle$ instead of $B(\langle\alpha\rangle)$, etc., for ease of readability.

that seem as innocent as these *must* fail to be provable? In Θ, as one would expect, all instances are provable. For $\alpha \in \Theta$ or $\alpha \notin \Theta$. In the second case, $\neg B \langle\alpha\rangle \in \Theta$, by (2). In either case, $\neg B \langle\alpha\rangle \vee \alpha \in \Theta$. Note, also, that all instances of the converse are also provable. For if $\alpha \in \Theta$, then $B \langle\alpha\rangle \in \Theta$, by (1), so $\neg\alpha \vee B \langle\alpha\rangle \in \Theta$. And if $\alpha \notin \Theta$, $\neg\alpha \in \Theta$, since Θ is complete; hence $\neg\alpha \vee B \langle\alpha\rangle \in \Theta$. In a sense then, B is a truth predicate since $B \langle\alpha\rangle \equiv \alpha \in \Theta$ (though this does not necessarily mean that $B \langle\alpha\rangle$ and α have the same truth values in the collapsed model).

2.5. The Naive Notion of Proof

We see, then, that inconsistent arithmetics can do a lot of nice things, and can avoid a number of features that many have held to be problematic for consistent arithmetic. This does not demonstrate that true arithmetic is inconsistent, but it certainly moves us in this direction. There are considerations that drive us further.

As is clear to anyone who is familiar with Gödel's theorem, at it's heart there lies a paradox. Informally, the "undecidable" sentence is the sentence 'This sentence is not provable'. Suppose that it is provable; then since whatever is proved is true, it is not provable. Hence, it is not provable. But we have just proved this. So it is provable after all (as well).[19] Let us look at this paradox more closely.

When mathematicians establish things to be true, they give proofs. These are informal deductive arguments, appealing to things which have already been proved or, ultimately, from things that are obviously true, and so where no proof is required. I will call the notion of proof in question here the *naive* notion of proof. Let us restrict ourselves to what can be proved naively about natural numbers. The language of naive proof about numbers is standard mathematical English (or some other natural language), but it is natural to suppose that this can be regimented into a suitable formal language, so that sentences may be assigned gödel-codes. Let us write $B_N(x)$ as a predicate of natural numbers which expresses the fact that x is (the code of) a sentence that is naively proved. B_N satisfies the following principles:

(3) $\vdash B_N \langle\alpha\rangle \supset \alpha$

(4) If $\vdash \alpha$ then $\vdash B_N \langle\alpha\rangle$

where $\vdash$ records naive proof. For (3): it is analytic that whatever is naively proved is true. Naive proof just is that sort of mathematical argument that

[19]The paradox is of the same kind as the "Knower paradox"; see Priest (1995), 10.5.

establishes something as true. And since this is analytic, it is itself naively provable. (Whether it is axiomatic or is derivable from more fundamental principles, we do not need to go into here). For (4): if something is naively proved then this fact itself constitutes a proof that α is proved.

But from these two principles, we can show that $\vdash$ is inconsistent. By usual methods of self-reference, we can construct a sentence that says of itself that it is not provable, i.e., a sentence, γ, of the form $\neg B_N \langle\gamma\rangle$. Substituting in (3) gives us $\vdash B_N \langle\gamma\rangle \supset \neg B_N \langle\gamma\rangle$, i.e., $\vdash \neg B_N \langle\gamma\rangle \vee \neg B_N \langle\gamma\rangle$. Thus, $\vdash \neg B_N \langle\gamma\rangle$; that is, $\vdash \gamma$. By (2), $\vdash B_N \langle\gamma\rangle$ (i.e., $\vdash \neg\gamma$). Arithmetic is therefore inconsistent.

I have not assumed that B_N is itself a predicate that can be constructed from the usual arithmetic vocabulary ($'$, $+$, $\times$, $=$). But there are, in fact, reasons to suppose that it can be.[20] It is part of the very notion of proof that a proof should be effectively recognised as such. For the very point of a proof is that it gives us a way of settling whether something is true. It is, therefore, a proof only when it is recognised as such. Thus, Dummett, for example, has stressed the point: it is part of the very notion of proof, unlike truth, that we can recognise one when we see it—at least in principle.[21]

Moreover, proof of the kind in question is a human practice. It is one that must be taught and learned. The human brain is, presumably, some sort of finite-state machine. It could not grasp the notion of proof if this were not axiomatic; if it were not, it would transcend the abilities of such a machine. For similar reasons, one must suppose that the grammar of any speakable language must be generated by a decidable set of rules. It might be pointed out that standards of proof may change over time, and that there is no reason to suppose that the change, itself, must happen in a rule-governed way. Indeed so. But we may take naive proof to comprise the standards of proof that are in operation here and now.

If naive proof in this sense is, indeed, axiomatic, then we can find a Σ_1 sentence of the standard language of arithmetic that expresses B_N. That is, arithmetic, as expressed in the usual vocabulary, is itself inconsistent. Nor is this technically unfeasible. In 2.4, we have already seen how a pure arithmetic can contain its own proof predicate and the attendant contradiction concerning its Gödel sentence.

---

[20]This argument is developed further in Priest (1987), ch. 3.

[21]See Dummett (1975).

2.6. Shapiro's Criticisms

In (2002) Shapiro, following ideas of Priest (1987), constructs an axiomatic theory, PA^*, that can prove its own Gödel sentence. The language of the theory contains a truth predicate, which is involved in the proof of the sentence;[22] but the Gödel sentence itself, as Shapiro emphasises, is purely arithmetic, employing only the proof predicate for the theory, which, being axiomatisable, is expressible in terms of $'$, $+$, $\times$ and $=$. Actually, the exact details of PA^* are left somewhat under-determined; but we need not go into that here. What I want to discuss are the unpalatable consequences that Shapiro supposes to follow from the fact that this theory can prove its own Gödel sentence. The features that Shapiro points to are possessed just as much by the axiomatisable inconsistent arithmetics that we looked at in section 2.4. I will therefore discuss his objections in the context of these.[23]

In the inconsistent arithmetic, Θ, both γ and $\neg\gamma$ are provable, where γ is purely arithmetic, and is of the form $\neg B\langle\gamma\rangle$. Since γ is provable, there is some number, g, which is the code of its proof. Hence, $Prov(\mathbf{g}, \langle\gamma\rangle)$ is true in the standard model, and so is provable in Θ. But $\neg B\langle\gamma\rangle$ is $\neg\exists y\, Prov(y, \langle\gamma\rangle)$, i.e., $\forall y \neg Prov(y, \langle\gamma\rangle)$. Hence $\neg Prov(\mathbf{g}, \langle\gamma\rangle)$ is provable as well. Now, $Prov(x, y)$ expresses a primitive recursive relation. Hence, if Θ is the true arithmetic, we have to accept that there are inconsistencies concerning numbers that are of this very basic kind. Worse, consider the following biconditionals. From left to right, they are unproblematic. Suppose that we accept them from right to left too.

P+ m is the code of a proof with of formula with code n iff $Prov(\mathbf{m}, \mathbf{n}) \in \Theta$

P− m is the not code of a proof with of formula with code n iff $\neg Prov(\mathbf{m}, \mathbf{n}) \in \Theta$

Then we have to accept that some number both is and is not the code of a

[22]The argument appeals to the claim that whatever is provable is true: $\forall x(B_N(x) \supset Tx)$ (plus the T-schema for the truth predicate). Given the meaning of the naive notion of proof, this is certainly analytically true, though it may well be provable from more fundamental things. There is, however, a way that the truth predicate can, in fact, be dispensed with. All that we actually need are the instances of the schema $B_N\langle\alpha\rangle \supset \alpha$, which are, equally, analytic. In fact, given the conditionals of this sort, the proof of the "undecidable sentence", $\gamma = \neg B_N\langle\gamma\rangle$, reduces to a very simple form. Substituting γ in the schema gives: $\neg B_N\langle\gamma\rangle \vee \gamma$. That is, $\gamma \vee \gamma$, i.e., γ.

[23]The material in this section arose from a seminar at the University of St Andrews at the end of 2002. I am grateful to the participants, and particularly to Steward Shapiro, for their helpful comments.

proof, and, more generally, that something could be both provable and not provable. What could this mean?

Shapiro offers three responses to this situation:

A Reject the soundness of Θ, on the basis of the fact that primitive recursive relationships are consistent.

B Accept that Θ is sound, but reject the biconditionals $\mathbf{P+}$ and $\mathbf{P-}$, and hence, on the assumption that Θ is the true arithmetic, the isomorphism between numbers with their operations and strings with theirs.

C Accept that Θ is sound, the biconditionals $\mathbf{P+}$ and $\mathbf{P-}$, and hence that something can both provable and not provable.

All of these options, Shapiro argues, should be resisted. If one is to take seriously the idea that Θ is the true arithmetic, option **A** is obviously not the way to go. One has to accept that even primitive recursive relations may be inconsistent. But this is not news. In the finite models of arithmetic even numerical equations can be inconsistent; that is, there can be truths of the form $\mathbf{m} = \mathbf{n} \wedge \mathbf{m} \neq \mathbf{n}$. One also has to accept, more generally, that even the computational part of mathematics is inconsistent. But this is not a problem either. Θ itself tells us exactly what an inconsistent computation theory is like. The Δ_0 formulas (that is, the sentences obtainable from equations using connectives and bounded quantifiers) express the recursive properties/relations.

Option **B** certainly involves jettisoning a connection in terms of which logicians have become accustomed to thinking. This is certainly a loss, though I do not think it as devastating as Shapiro does. However, it seems to me that the simplest and most natural response is option **C**, so I will discuss this option at length. Shapiro marshals essentially two considerations against it. Let us consider these in turn.

2.7. The Inconsistency of Peano Arithmetic

Shapiro's first objection, and the quicker to deal with, is that if one holds that primitive recursive relations are inconsistent, it follows not just that Θ is inconsistent, but that Peano Arithmetic (PA) is inconsistent—which seems implausible. The reason is that all recursive relationships are known to be representable in PA.

The reply is simply that if the recursive relationships are as specified by Θ, they are *not* all representable in PA—just because it is consistent. Where does the proof of the fact that all recursive relationships are representable in PA break down, however? The answer depends on which proof we are talking about, and on which inconsistent theory of arithmetic is correct. But let us suppose, for the sake of argument, that Θ is $Th(\mathcal{M}^6_{10})$ (refer to the diagram of 1.4); and let us look at a direct proof to the effect that the formula $x = y'$ represents the successor relation in Θ.[24] We need to show that if $i = j'$ then $\mathbf{i} = \mathbf{j}' \in PA$. This is proved by induction on j. Suppose that $j = 0$. Then if $i = 0'$, $i = 1$, and $\mathbf{1} = \mathbf{0}' \in PA$. Now suppose that the result holds for j, and show it for j'. So suppose that $i = j''$. Since i is not 0, there is a k such that $i = k'$. Hence, $k' = j''$, and $k = j'$. By induction, $\mathbf{k} = \mathbf{j}' \in PA$; so $\mathbf{i} = \mathbf{k}' = \mathbf{j}'' \in PA$.

The second part of this argument breaks down for $Th(\mathcal{M}^6_{10})$, since a number may have multiple predecessors, some of them *greater* than itself. Thus, suppose that j is 8. If $i = 8''$ then certainly for some k, $k' = 8''$; k can be 9 or 15. Now, $9' = 8''$, $9 = 8'$, $\mathbf{9} = \mathbf{8}' \in PA$ (by induction), and so $\mathbf{9}' = \mathbf{8}'' \in PA$. But, though $15' = 8''$, it does not follow that $15 = 8'$, so the argument breaks down. Indeed, $\mathbf{15} = \mathbf{8}' \notin PA$.

Thus, and in general, if you take Θ to provide the correct account of recursive relationships, then these will be representable (trivially) in Θ; but PA will be incomplete, since it captures only a consistent fragment of the truth. Dually, of course, if you take the usual classical line on recursive relationships, PA will be complete, but Θ will give *more* than the truth, because it is inconsistent. In other words, if you match up the formal arithmetic and the theory of recursive relations properly, then you will get representability. But if you mis-match these by taking one to be consistent and the other not, then things will go wrong.

2.8. The Incredulous Stare

Shapiro's other main objection amounts to a version of the incredulous stare. Let me put the start of it in his own words:[25]

> On all accounts—including the non-dialetheic perspective—we have that g is the code of a Θ-derivation of γ. This can be verified with a painstaking, but completely effective check. How can the dialetheist go on to maintain that, in addition, g is *not* the code of a Θ-derivation of γ? What does it mean to say this?

[24] Proofs of this kind can be found in Boolos and Jeffrey (1974), ch. 14, Part III.

[25] Shapiro (2002), p. 828. I have changed the notation to bring it in line with that used in this essay. The italics are original.

> Since $\neg Prov$ is recursive predicate, we can supposedly verify—at the same time, in almost exactly the same way—that g is not the code of a Θ-derivation of γ. How?

Shapiro asks how we can possibly verify a sentence expressing a recursive relation and its negation? What can this mean?

In principle, the answer is easy. Since we are endorsing **P**+ and **P**−, we are now taking seriously the thought that metatheoretic sentences may be contradictory. If so, they must play by the same rules as those of Θ, and in particular, be based on the logic LP (or FDE). In any theory based on LP or FDE, α and $\neg\alpha$ are verified by *different* procedures. Thus, e.g., to determine whether $t_1 = t_2$ is true, we have to look to see whether $\langle d(t_1), d(t_2)\rangle \in E_=$. To determine whether $t_1 \neq t_2$ is true, we have to look to see whether $\langle d(t_1), d(t_2)\rangle \in A_=$. These are separate matters. Thus, in $Th(\mathcal{M}^6_{10})$, once we have checked to see whether $\mathbf{i} = \mathbf{j}$, the question of whether $\mathbf{i} \neq \mathbf{j}$ is a *further* question. $\mathbf{0} = \mathbf{0}$ is true, but $\mathbf{0} \neq \mathbf{0}$ is not. $\mathbf{10} = \mathbf{10}$ is true, but so is $\mathbf{10} \neq \mathbf{10}$.

Thus, to bring the matter to bear on proof explicitly, suppose that g is the code of a proof of γ. Suppose, for the sake of argument, that the code is 37. Then to say that g is the code is to say something equivalent to $g = 37$. What does it mean to say that it is also not the code of a proof of γ? It is to say that $g \neq 37$ as well. This is the case if $37 = 37 \wedge 37 \neq 37$, which it can be in an inconsistent arithemtic.

And what does it mean to say that γ is both provable and not? To say that it is provable is to say that $\exists x(x$ is the code of a proof of $\gamma)$, i.e., on the supposition at hand, $\exists x\ x = 37$. To say that it is not provable is to say that $\neg\exists x(x$ is the code of a proof of $\gamma)$ i.e., $\forall x\neg(x$ is the code of a proof of $\gamma)$, i.e., $\forall x\neg(x = 37)$, i.e., $0 \neq 37 \wedge 1 \neq 37 \wedge ... \wedge 37 \neq 37 \wedge ...$; which is, of course, true if $37 \neq 37$. In other words, to say that γ is not provable is to say that every number is distinct from a code of the proof of γ. This does not rule out there being a proof of γ. (In general, the truth of $\neg\alpha$ in a paraconsistent setting does not rule out the truth of α.) In particular, it will hold if the proof is distinct from itself. And how can a proof be distinct from itself? In the same way that a number can. After all, on option **C**, the one at issue, we are retaining the structural identity between strings and numbers. Both are, after all, abstract objects. And the inconsistent behaviour of strings is just as good or bad as the inconsistent behaviour of numbers.[26]

[26] It is worth noting that if numbers have inconsistent properties, then this will affect their behaviour whatever theory they are taken to be coding. Thus, the gödel codes

There are, or course, concrete objects whose behavior in some sense represents the behaviour of the abstract objects. In this case, there are marks of dried ink on paper that represent the abstract strings. But no one expects the properties of the abstract objects to carry over of necessity to their physical representations. Thus, $3 > 1$; but a 9-point token of '3' is much smaller than a 18-point token of '1'. So the mere fact that, e.g., $37 \neq 37$ does not necessarily mean that we have to find some concrete object that is not self-identical. Of course, we need not rule out this possibility either. Talk of concrete objects that are not self-identical immediately takes one's thought down the path of sub-atomic particles and their curious behaviour. But this is not the place to follows such speculations.

2.9. Conclusion

Shapiro's objections stem from being half-hearted about dialetheism. If one endorses an inconsistent arithmetic, but tries to hang on to either a consistent computational theory or a consistent metamathematics of proof, one is in for trouble. The solution to his problems is therefore not to be half-hearted, and to accept that these other things are inconsistent too. Indeed, the arithmetic itself shows us how to do this: the facts about computability and provability are simply read off from the arithmetic.[27]

Discussions in the philosophy of mathematics are always built on shaky foundations if they are not underpinned with the appropriate technical material. This is certainly true of discussions of the inconsistency of arithmetic. The inconsistent models show us exactly what can be done and how. That

of PA will behave just as inconsistenly as those of Θ. In other words, if numbers are inconsistent, we may expect things to be both provable and not provable in PA just as much as in Θ. This does not, of course, mean that PA is itself inconsistent. As to where the inconsistency of gödel codes arises, it might be only for numbers so large that they are larger than anything that is humanly meaningful. (See Priest (1994a).)

[27]The philosophical discussion has appealed to various metatheoretic properties of inconsistent arithmetics. How were these established? A natural assumption is that they were proved in a classical (consistent) metatheory, such as ZF. If we are now endorsing an inconsistent (meta-)arithmetic, we can no longer be working in ZF. What entitles us to be sure that we may still invoke those results? One answer goes essentially as follows. Start with a model of ZF, say (an initial segment of) the cumulative hierarchy. Then use the Collapsing Lemma to produce a collapsed model of ZF in which the structure of the numbers brings it into line with the inconsistent arithmetic we are envisaging. (For collapsed models of ZF, see Priest (1995), Part 3, Technical Appendix.) We can take the theory of that collapsed model to provide the metatheory in which we are working. And just as any theorem of standard arithmetic holds in the theory of a collapsed model of arithmetic, so any theorem of ZF holds in that theory.

hardly settles many of the interesting philosophical questions. But it does put a firm skeleton below the philosophical flesh.

References

[1] Batens, D. (200a), 'The Demise of Rich Finitism; a Study in the Limitations of Paraconsistency' to appear.

[2] Boolos, G. and Jeffrey, R. (1974), *Computability and Logic*, Cambridge: Cambridge University Press.

[3] Chihara, C. (1984), 'Priest, the Liar and Gödel', *Journal of Philosophical Logic* 13, 117-24.

[4] Denyer, N. (1995), 'Priest's Paraconsistent Arithmetic', *Mind* 104, 567-75.

[5] Dummett, M. (1975), 'The Philosophical Basis of Intuitionist Logic', pp. 5-40 of H. Rose and J. Shepherdson (eds.), *Logic Colloquium '73*, Amsterdam: North Holland; reprinted as ch. 14 of Dummett's *Truth and Other Enigmas*, London: Duckworth.

[6] Dunn, J. M. (1979), 'A Theorem of 3-Valued Model Theory with Connections to Number Theory, Type Theory and Relevance', *Studia Logica* 38, 149-69.

[7] Gasking, D. A. T. (1940), 'Mathematics and the World', *Australasian Journal of Philosophy* 18, 97-116; reprinted as pp. 390-403 of Benacerraf, P. and Putnam, H. (eds.), *Philosophy of Mathematics; Selected Readings*, Oxford: Oxford University Press, 1964.

[8] Kaye, R. (1991), *Models of Peano Arithmetic*, Oxford: Clarendon Press.

[9] Meyer, R. K. (1976), 'Relevant Arithmetic', *Bulletin of the Section of Logic, Polish Academy of Sciences* 5, 133-7.

[10] Meyer, R. K. and Mortensen, C. (1984), 'Inconsistent Models for Relevant Arithmetics', *Journal of Symbolic Logic* 49, 917-29.

[11] Mortensen, C. (1987), 'Inconsistent Nonstandard Arithmetic', *Journal of Symbolic Logic* 52, 512-8.

[12] Mortensen, C. (1988), 'Inconsistent Number Systems', *Notre Dame Journal of Formal Logic* 29, 45-60.

[13] Mortensen, C. (1990), 'Models for Inconsistent and Incomplete Differential Calculus', *Notre Dame Journal of Formal Logic* 31, 274-85.

[14] Mortensen, C. (1995), *Inconsistent Mathematics*, Dordrecht: Kluwer Academic Publishers.

[15] Mortensen, C. and Meyer, R. K. (1985), 'Relevant Quantum Arithmetic', pp. 211-26 of L. De Alcantara (ed.), *Mathematical Logic and Formal Systems*, New York, NY: Marcel Dekker.

[16] Nelson, D. (1959), 'Negation and Separation of Concepts in Constructive Systems', pp. 208-25 of A. Heyting (ed.), *Constructivity in Mathematics*, Amsterdam: North Holland Publishers.

[17] Priest, G. (1979), 'Logic of Paradox', *Journal of Philosophical Logic* 8, 219-41.

[18] Priest, G. (1987), *In Contradiction*, Dordrecht: Kluwer Academic Publishers.

[19] Priest, G. (1991), 'Minimally Inconsistent *LP*', *Studia Logica* 50, 321-31.

[20] Priest, G. (1994a), 'Is Arithmetic Consistent?', *Mind* 103, 321-31.

[21] Priest, G. (1994b), 'What Could the Least Inconsistent Number be?', *Logique et Analyse* 37, 3-12.

[22] Priest, G. (1995), *Beyond the Limits of Thought*, Cambridge: Cambridge University Press; second and revised edition, Oxford: Oxford University Press, 2002.

[23] Priest, G. (1996a), 'On Inconsistent Arithmetics: Reply to Denyer', *Mind* 105, 649-59.

[24] Priest, G. (1996b), 'On Alternative Geometries, Arithmetics and Logics: a Tribute to Łukasiewicz', a paper given at the conference *Łukasiewicz in Dublin*; now forthcoming in *Studia Logica*.

[25] Priest, G. (1997), 'Inconsistent Models of Arithmetic; I Finite Models', *Journal of Philosophical Logic* 26, 223-35.

[26] Priest, G. (1998), 'What's so Bad about Contradictions?', *Journal of Philosophy* 95, 410-26.

[27] Priest, G. (2000), 'Inconsistent Models of Arithmetic; Part II, the General Case', *Journal of Symbolic Logic* 65, 1519-29.

[28] Priest, G. (2001), *Introduction to Non-Classical Logic*, Cambridge: Cambridge University Press.

[29] Shapiro, S. (2002), 'Incompleteness and Inconsistency', *Mind* 111, 817-32.

[30] Van Bendegem, J.-P. (1993), 'Strict, Yet Rich Finitism', pp. 61-79 of Z. W. Wolkowski (ed.), *First International Symposium on Gödel's Theorems*, Singapore: World Scientific Press.

[31] Van Bendegem, J.-P. (1994), 'Strict Finitism as a Viable Alternative in the Foundations of Mathematics', *Logique et Analyse* 37, 23-40.