Hopes Fade For Saving Truth

GRAHAM PRIEST

Critical Notice: Hartry Field, *Saving Truth from Paradox*; Oxford: Oxford University Press, 2008; pp. xv + 406; ISBN 978-0-19-923075-4, £55 (hbk); ISBN 978-0-19-923074-7, £19.99 (pbk).

1. Introduction

In a series of about eight papers, starting in 2002, Hartry Field has been developing a new and distinctive consistent solution to the semantic paradoxes. The present book brings together this work in a uniform manner. It does more than this, however. It surveys the current lie of the land, and assesses alternative views – from Tarski to dialetheism. The book is technically deft, philosophically shrewd and insightful, and written with a wry sense of humour. It is essential reading for anyone interested in the area: whether or not the reader ends up agreeing with Field, they will come away much wiser. It is, as the very first sentence of the preface says, 'opinionated'; views are sometimes dismissed rather summarily.[1] And I doubt that this will be a definitive statement of Field's views: the book leaves a number of loose ends hanging, and Field is the sort of philosopher who is always revising his views in the light of new thoughts. Nonetheless, in my opinion, the book is the most significant publication on consistent solutions to the liar paradox since Kripke's seminal 'Outline of a Theory of Truth'.[2]

2. General overview

After an introduction, the book falls into five parts. The first discusses a number of background issues for what is to be covered: Tarski's indefinability theorem and its import; the connection between model-theoretic validity and truth preservation; questions

[1] To give just two examples from the first couple of pages of the book: 'None of this seems terribly attractive' (7), 'which I find very hard to get my head around' (10–11). All page and section references are to Field's book, unless otherwise specified. Italics in all quotations are original.

[2] S. Kripke, 'Outline of a Theory of Truth', *Journal of Philosophy* **72** (1975), 690–716.

doi:10.1017/S0031819109990489

concerning the ability of a theory to prove itself to be sound; Kripke's solution to the paradoxes (of which Field's view is, in some sense, a development); Łukasiewicz' continuum-valued semantics, and especially the behaviour of its conditional (closely related to Field's own conditional); possible connections between semantic paradoxes, vagueness, and definability paradoxes (such as Berry's).

The second part of the book surveys theories dubbed 'broadly classical'. These are theories based on classical logic, but whose language contains its own truth-predicate. Split the T-schema into two:[3]

Capture: $A \rightarrow T\langle A\rangle$
Release: $T\langle A\rangle \rightarrow A$

Chapter 7 deals with theories which endorse Release, but only restricted versions of Capture. These include the Feferman 'closing off' of a Kripke fixed-point model (*KF*). Chapter 8 deals with theories which endorse Capture, but only restricted versions of Release. These have problems that are, generally speaking, duals to those in Chapter 7. The next four chapters discuss theories which endorse neither Capture nor Release, but only their restricted versions, $A \vDash T\langle A\rangle$ and $T\langle A\rangle \vDash A$. The theories also reject the principle:

$$\frac{A \vdash C \quad B \vdash C}{A \vee B \vdash C}$$

Accounts which supervaluate on a partially-defined truth predicate are of this kind, as are theories of the 'revision-theoretic' kind, such as those of Gupta and Herzberger.[4] Field argues that these have a most important failing: $T\langle A\rangle$ and A are not logically intersubstitutable (Transparency). The final chapter in this part discusses views which 'fragment' the truth predicate, such as Tarski's and contextualist approaches.

The chapters so far provide a rich survey of part of the contemporary landscape. But they also lay the ground for an exposition of Field's own view, which is contained in the next two parts of the book. (In fact,

[3] I use T for the truth predicate, and angle brackets for an appropriate naming device. Warning: I sometimes use notation different from Field's in the cause of simplicity.

[4] Interestingly, the dual subvaluational theories which reject:

$$\frac{A \vdash B \quad A \vdash C}{A \vdash B \wedge C}$$

go under Field's radar.

readers of these parts of the book will need to go back to earlier chapters sometimes to grasp the full import of Field's discussion of his own positive account.) Views with which Field is sympathetic are labelled 'paracomplete', which means that the Law of Excluded Middle (LEM) fails: $\nvDash A \vee \neg A$. What Field seeks is a logic of this kind, with an appropriate conditional, and which accepts Transparency. Chapter 16 considers a couple of ways in which one might try to produce such a logic, and finally settles on Field's preferred view. The specification of the logic is model-theoretic, and an appropriate conditional is defined by induction over the ordinals, using a revision-theoretic strategy. Chapter 17 converts this into an algebraic semantics, where values are certain functions from ordinals to $\{0,i,1\}$, and gives readers a feel for which inferences are valid, and which are not. (An appropriate proof-theory is not given; nor can it be since, as Field explains, the logic is not axiomatisable.)

Part four of the book contains a discussion of a variety of issues which might be thought to constitute objections to the account. Chapters 20 and 21 deal with Curry's paradox, the failure of the logic to declare its own inferences truth-preserving, and various other paradoxes, notably definability paradoxes. The next three chapters deal with the question of whether the solution succeeds only because the language is expressively incomplete, and the intimately related question of so called 'revenge paradoxes'. Boolean negation is dealt with in Chapter 21; determinate truth of various kinds in Chapter 22; and Chapter 23 mops up, including some things which Field takes to be 'genuine costs' of his approach (none of which, unsurprisingly, he finds to be very serious).

The final part of the book is a discussion of a dialetheic approach to the paradoxes. Chapter 24 explains essentially what this is; Chapter 25 provides some warm-up objections. The next two chapters concern the ability of dialetheic theories to endorse the truth-preservingness of their own inferences, their ability to prove themselves sound, and the question of expressive incompleteness and revenge paradoxes. On all counts, Field argues, his approach is at least as good; and on most, better.

3. Field's account

There is much in this book with which I agree, and much more which is worthy of discussion; but in what follows I will concentrate on Field's own account and his comparison of this with a dialetheic approach.

Let us start with a closer look at the details of Field's semantics. I restrict myself to the propositional part of his logic. The semantic behaviour of quantifiers is orthodox.

What Field is after is a solution to the semantic paradoxes with the following features:

1. The language contains a truth predicate and the language of arithmetic – or, more generally, set theory.
2. For all A, A and $T\langle A\rangle$ are logically inter-substitutable. (Transparency.)
3. There is a model in which the natural numbers behave standardly. (The failure of Łukasiewicz' continuum-valued semantics to provide this is the main reason for rejecting it.)
4. The theory is consistent.
5. The theory has an appropriate conditional.

The cornerstone of achieving these things is the rejection of LEM.

The specification of the theory, as described in Chapter 16, piggybacks on Kripke's fixed-point construction. For this, take a first-order language (with conditional $\supset$). Call this the *ground language*. Augment it with a monadic truth predicate, T, and a binary connective, $\rightarrow$. For the purpose of this part of the construction we take formulas of the form $A \rightarrow B$ to be atomic, as well as those given by T and the predicates of the ground language. An interpretation of the language is that of the Strong Kleene logic, K_3.[5] There are three values, ordered as follows: $0 < i < 1$. Take a ground model, M_0, with evaluation function, ν_0. The interpretations of all the predicates in the ground language are classical (that is, no atomic sentence involving them takes the value i), and the interpretation of the arithmetic machinery is standard. This delivers an appropriate gödel coding of formulas. Let $\langle A\rangle$ be the numeral of the gödel code of A. For all A and B, $\nu_0\,(A \rightarrow B) = i$, and for any closed A, $\nu_0\,(T\langle A\rangle) = i$. (These are the simplest options; there are others.) We now define a sequence of interpretations, M_α (with interpretation functions ν_α), by recursion over the ordinals, as follows. The only thing that changes as we ascend the ordinals is the value of $T\langle A\rangle$ for closed A. Given an ordinal, γ, say that a condition, $\Theta(\beta)$ *eventually holds by* γ iff $(\exists\alpha < \gamma)\forall\beta$ (if $\alpha \leq \beta < \gamma$ then $\Theta(\beta)$). (If γ is a successor ordinal, $\Theta(\beta)$ eventually holds by γ iff $\Theta(\beta - 1)$.) Then:

$\nu_\gamma\,(T\langle A\rangle) = 1$ if $\nu_\beta\,(A) = 1$ eventually holds by γ
$\nu_\gamma\,(T\langle A\rangle) = 0$ if $\nu_\beta\,(A) = 0$ eventually holds by γ
$\nu_\gamma\,(T\langle A\rangle) = i$ otherwise

We can establish that there is a δ, such that $\nu_\delta = \nu_{\delta+1}$ (the "Fixed-Point" Theorem). Let $M_0^* = M_\delta$ for the least such δ. The

[5] See G. Priest, *Introduction to Non-Classical Logic: from If to Is* (Cambridge: Cambridge University Press, 2008), 7.3.

theory of M_0^* satisfies all of desiderata 1–4, but the logic of the conditional is clearly too arbitrary. For example, instances of $A \leftrightarrow T\langle A\rangle$ are not true (that is, do not take value 1) at M_0^*. (This can be rectified by a different choice of M_0, but the general point remains.)

To obtain an appropriate conditional, we iterate the fixed point construction itself through the ordinals. Now as we ascend the ordinals, the things that change are the values of formulas of the form $A \rightarrow B$, providing new ground models for the Kripke construction at each stage. Specifically:

$\nu_\gamma(A \rightarrow B) = 1$ if $\nu^*_\beta(A) \leq \nu^*_\beta(B)$ eventually holds by γ
$\nu_\gamma(A \rightarrow B) = 0$ if $\nu^*_\beta(A) > \nu^*_\beta(B)$ eventually holds by γ
$\nu_\gamma(A \rightarrow B) = i$ otherwise

This generates a sequence of corresponding fixed points, $M_0^*, M_1^*, \ldots$. These do not, themselves, reach a fixed point, but there is a certain kind of stability. Define the *ultimate value* of a formula A, $|A|$, as follows:

$|A| = 1$ if $\exists\alpha(\forall\beta \geq \alpha)\ \nu^*_\beta(A) = 1$
$|A| = 0$ if $\exists\alpha(\forall\beta \geq \alpha)\ \nu^*_\beta(A) = 0$
$|A| = i$ otherwise

Field is then able to prove his 'Fundamental Theorem'. Call an ordinal, β, *acceptable* iff, for every A, $\nu^*_\beta(A) = |A|$. Then the Theorem is that: $\forall\alpha(\exists\beta > \alpha)$ β is acceptable. An inference is valid if it preserves ultimate value 1 with respect to every ground model.[6] The construction now satisfies all of desiderata 1–5.

4. Field's Logic and Relevant Logic

In Chapter 17, Field reworks his semantics into an algebraic semantics. Given a ground model, M_0, let δ be the least acceptable ordinal, and let $\delta + \sigma$ be another acceptable ordinal. The values of the algebra are members of a certain subset of the functions from σ to $\{1, i, 0\}$. An ordering is defined point-wise: $f \preceq g$ iff for all $\alpha \in \sigma, f(\alpha) \leq g(\alpha)$. The result is a De Morgan algebra with maximum element **1** (where for all $\alpha \in \sigma$, $\mathbf{1}(\alpha) = 1$). Validity is defined in terms of preservation of the value **1** in all the algebras generated by all ground models. Given a suitable choice of the subset of functions (the details of which are

[6] In fact, Field gives a definition of validity only when he reformulates the semantics as an algebraic one in the next chapter. But this is what it is.

somewhat complex, but need not concern us here), the original and algebraic notions of validity coincide.

The semantics of Chapter 16 are probably easier to work with than the algebraic semantics, but the reformulation does bring out some interesting points. In particular, the algebraic structures in question are De Morgan algebras. De Morgan algebras are the algebraic structures behind relevant logics.[7] One should therefore expect a close connection between Field's logic and relevant logics. From the algebraic perspective, the main difference is that in relevant logics the algebra contains a designated element, e, satisfying certain properties; and $a \preceq b$ iff $e \preceq a \rightarrow b$. Field has the special case in which e is always $\mathbf{1}$.

Field's logic contains the relevant logic *DW*, and shares many of its properties.[8] In particular, it rejects the LEM:

$$\vDash A \vee \neg A$$

Permutation:

$$A \rightarrow (B \rightarrow C) \vDash B \rightarrow (A \rightarrow C)$$

and two forms of Contraction:[9]

$$A \rightarrow (A \rightarrow B) \vDash A \rightarrow B$$
$$\vDash (A \wedge (A \rightarrow B)) \rightarrow B$$

Contraction is very important in connection with Curry paradoxes, as we shall see in due course. And since the logic validates *modus ponens* (MP), $A \wedge (A \rightarrow B) \vDash B$, 'conditional proof' (CP) breaks down. That is, it is not the case that:

if $C \vDash D$ then $C \rightarrow D$.

It is worth noting that the relevant logic *DW* also satisfies desiderata 1–4 and, arguably, 5. (Though Field contests 5, as we shall see in due course.) This follows, essentially, from a result of Brady.[10]

[7] See G. Priest, 'Paraconsistent Logic', 287–393, Vol. 6, 2nd edn. of D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic* (Dordrecht: Kluwer Academic Publishers, 2001), 5.6.

[8] For *DW*, see Priest, *Introduction to Non-Classical Logic*, ch. 10.

[9] Given some simple and relatively uncontentious assumptions, the latter entails the former. For suppose that $(A \wedge (A \rightarrow B)) \rightarrow B$. Then $(A \rightarrow (A \wedge (A \rightarrow B))) \rightarrow (A \rightarrow B)$. But given $A \rightarrow (A \rightarrow B)$ then $A \rightarrow (A \wedge (A \rightarrow B))$, since $A \rightarrow A$. So we have $A \rightarrow B$.

[10] R. Brady, 'The Simple Consistency of Set Theory Based on the Logic *CSQ*', *Notre Dame Journal of Formal Logic* **24** (1983), 431–439. See Priest, 'Paraconsistent Logic', 8.2. Brady's proof uses a double recursion of the same kind employed by Field. As far as I know, Brady was the first

Field's logic is, however, not a relevant logic, since we have all of the following:

$\models (A \wedge \neg A) \rightarrow (B \vee \neg B)$
$\neg(A \rightarrow B) \models (B \rightarrow A)$
$A \vee \neg A, B \vee \neg B \models (A \supset B) \leftrightarrow (A \rightarrow B)$

The last of these means that in the presence of the relevant instances of LEM Field's conditional collapses into the classical material conditional, and so we may reason classically with it. Because of the way that ground models work, this means that any T-free sentence has ultimate value 1 or 0, and so we may reason with such sentences using classical logic. In particular, if the language contains the set-membership predicate, ε, and the interpretation of the ground language is a model of ZF, all the (T-free) theorems of ZF are going to have ultimate value 1. Thus, we may take Field's model-theoretic construction, as he does, to be carried out in ZF. (In fact, with the relevant instances of LEM, the $\supset$ of K_3 collapses into the $\supset$ of classical logic, and so we can reason classically employing sentences in the ground language anyway.)

5. Truth Preservation

The standard liar paradox, 'This sentence is false', can be handled by rejecting the LEM, as Field does. But Curry's paradox concerns only truth and the conditional essentially. Hence, it is of great concern to Field. The paradox involves a sentence, K, of the form $T\langle K\rangle \rightarrow \perp$ (where $\perp$ is a logical constant that entails everything – or just any proposition one likes). Given this, we may reason as follows:

Assume:	$T\langle K\rangle$	(1)
Release, MP:	K	
i.e.:	$T\langle K\rangle \rightarrow \perp$	
MP:	$\perp$	
CP:	$T\langle K\rangle \rightarrow \perp$	discharging (1)
i.e.:	K	
Capture, MP:	$T\langle K\rangle$	
MP:	$\perp$	

person to show that it was possible to apply persistence methods of this kind to non-monotonic connectives.

What this shows is that, assuming that there is a self-referential sentence of the form K, *no one* can accept all of MP, CP, Release, and Capture. Field's semantics invalidates (just) CP. (It is worth noting that this move is quite independent of the LEM. The LEM fails in intuitionist logic, but CP holds. Conversely, if one adds the LEM to DW, CP still fails.)

There is, as Field stresses, a close connection between Curry paradoxes and truth-preservation. Call an inference, $A \vdash B$, *Truth Preserving* if the conditional $T\langle A\rangle \rightarrow T\langle B\rangle$ holds.[11] It is natural to suppose that valid inferences are truth preserving in this sense. The obvious justification goes as follows:

Assume:	$A \vDash B$	
Assume:	$T\langle A\rangle$	(1)
Release, MP:	A	
So:	B	
Capture, MP:	$T\langle B\rangle$	
CP:	$T\langle A\rangle \rightarrow T\langle B\rangle$	discharging (1)

This works for no one, since it uses all four of the Curry-generating principles. Moreover, if we have the validity of MP, Truth Preservation delivers $T\langle A\wedge(A \rightarrow B)\rangle \rightarrow T\langle B\rangle$. Given the T-schema, $(A\wedge(A \rightarrow B)) \rightarrow B$, so we have Curry's paradox. Thus, given MP and the T-schema, one cannot have Truth Preservation. One could, of course, maintain Truth Preservation by rejecting the T-schema; but, without this, it is not clear why anyone would be interested in having Truth Preservation. Certainly, then, the failure of Truth Preservation cannot be held against Field's account, or any other.

The failure of Truth Preservation does, however, have other significant consequences. The intuitive proof of the soundness of an axiomatic theory, say of arithmetic, is an inductive argument, starting from the truth of the axioms, and using the fact that the rules of inference are truth preserving. Suppose that one has an axiomatic theory using MP, in a language with its own truth predicate. If we try to carry out the proof of soundness in the theory itself, the obvious method for proving the inductive step for MP requires Truth Preservation. Thus, suppose that *Prov* is the proof predicate for the theory. Soundness is a statement of the form $\forall x(Prov(x) \rightarrow Tx)$. Now suppose that we have proved soundness for deductions of

[11] As we shall see in due course, truth-preservation can be said in many ways. Hence the capitals to mark this particular form.

length $\leq n$, and that our deduction of length $n+1$ uses an instance of MP from x and $x \rightarrow y$.[12] Then by induction hypothesis, we have $Prov(y) \rightarrow T(x)$ and $Prov(x \rightarrow y) \rightarrow T(x \rightarrow y)$. To finish the job, we need (at least) $T(x \wedge (x \rightarrow y)) \rightarrow Ty$, and if we have this in a universally quantified form, we have every instance of the Curry-generating $T\langle A \wedge (A \rightarrow B)\rangle \rightarrow T\langle B\rangle$. Hence, the proof cannot go through.

Of course, the breakdown of a soundness proof is no surprise in the case of classically based theories, because of Gödel's Second Incompleteness Theorem. But one might have hoped that a non-classical theory could fare better. Nor is it just that the obvious proof of soundness breaks down. Any consistent theory, Field's included, which endorses the T-schema and uses classical logic to reason about arithmetic cannot endorse soundness at all. Suppose it did. Let G be the Gödel sentence of a theory with its own truth predicate, $\neg Prov\langle G\rangle$. Then we have $Prov\langle G\rangle \rightarrow T\langle G\rangle$, $T\langle G\rangle \rightarrow G$ $(=\neg Prov\langle G\rangle)$, and so $Prov\langle G\rangle \rightarrow \neg Prov\langle G\rangle$. By classical reasoning, we have proved $\neg Prov\langle G\rangle$, that is, G, and so $Prov\langle G\rangle$. Now, there is something very strange about a theory that is not able to endorse its own soundness. If we get the proof-theory right, *of course* it is true that it is sound. How can the mere fact that the proof theory contains this truth make it impossible?

6. Other Paradoxes

Liar and Curry paradoxes aside, there is a host of other self-referential paradoxes that a comprehensive solution must deal with. For a start, the argument concerning G which we met at the end of the last section, is 'Gödel's paradox'. The upshot of this paradox for Field is orthodox: we simply cannot accept that $Prov\langle G\rangle \rightarrow G$.[13] We will come back to this later. But there is another paradox (not explicitly mentioned by Field), which is structurally similar, and which also causes trouble; this is the Knower paradox. Let K be the predicate 'is known'. This is not in Field's language, but the insistence that it must be kept out would be as arbitrary as the insistence that the truth predicate for a language must be kept out of it. Now, consider a sentence, H, of the form $\neg K\langle H\rangle$ ('This sentence is not known'). We have $K\langle H\rangle \rightarrow H$ $(=\neg K\langle H\rangle)$. An appeal to LEM, establishes

12 By which I mean $C(x,y)$, where this refers to the sentence obtained by inserting ' $\rightarrow$ ' between x and y.

13 Löb's theorem tells us that $Prov\langle A\rangle \rightarrow A$ is provable iff A is.

that $\neg K\langle H\rangle$ (=H), so $K\langle H\rangle$. Given the veridicality of knowledge, it is hard to reject $K\langle H\rangle \to H$. Field, I am sure, would insist that K, like T, is one of those predicates that generates 'indeterminacies', and so for which we cannot assume the LEM.

The trouble with this strategy is that belief, and so knowledge, is a contingent matter. Let us suppose that it is Field's own knowledge that is at issue. (That is, we interpret K as 'Field knows that'.) Field rejects $K\langle H\rangle \vee \neg K\langle H\rangle$. So he rejects both H and $\neg H$. In particular, then, he does not believe H; so he does not know H: $\neg K\langle H\rangle$. So $K\langle H\rangle \vee \neg K\langle H\rangle$, and we have the LEM.[14]

Other paradoxes which cannot be ignored (which Field does consider) are those in the family of definability paradoxes, such as Berry's and König's. Schematically, these go as follows. Taking ourselves to be quantifying over well-ordered numbers of some kind, let Dxy be 'x denotes y (in less than 100 words, etc.)'. By the appropriate cardinality reasoning, we can prove $\exists y \neg \exists x Dxy$. Let μ be a least-number operator, and let m be $\mu y \neg \exists x Dxy$. By the schema, M:

$$\exists y A(y) \vdash A(\mu y A(y))$$

we have $\neg \exists x Dxm$. But $D\langle m\rangle m$ (the naive D-schema, which follows from the naive truth-of schema). So $\exists x Dxm$.

According to Field, such arguments fail because M fails. What holds is M':

$$\exists y(A(y) \wedge (\forall x < y)(A(x) \vee \neg A(x)) \vdash A(\mu' y A(y))$$

where $<$ is the ordering of the numbers in question. The claim that M' holds is justified by defining $\mu' y A(y)$ in terms of a definite description operator, ι. The term $\iota y A(y)$ is to refer to the unique y satisfying $A(y)$. (If there is no such thing, how it behaves is another matter, and not relevant here.) We then define $\mu' y A(y)$ as $\iota y(A(y) \wedge (\forall x < y) \neg A(x))$. In other words, $\mu' y A(y)$ refers to the least thing such that it satisfies $A(y)$ and all smaller things satisfy $\neg A(y)$. The validity of M' then follows.

But all this is beside the point. It is no reply to an argument that a certain principle generates a contradiction to point out that a different principle does not. We need to know about M itself. Of course, if you define μ as μ', M is not going to hold. But why define it like that? – in fact, why define it at all? According to Field's semantics, for any

[14] Here is another way to put the argument. Field rejects H; he does not believe this, and so does not know it, $\neg K\langle H\rangle$; that is, H. By introspection, he knows this, so $K\langle H\rangle$.

condition $A(y)$, if there is anything which, when assigned to y, gives the value 1, there is a least such. All the lesser numbers may give the value 0 or the value i. That does not matter. (Recall that the semantics is carried out in *ZF*, and so we may reason classically.) Let $\mu y A(y)$ denote *that*. This would appear to be a perfectly intelligible least-number operator (distinct from μ'), and one which satisfies M.

In an earlier publication,[15] Field argued that the addition of an operator satisfying M must be rejected, on the ground that it could be used to establish the LEM generally; but as I pointed out in reply,[16] the argument fails, since it just assumes that $1 = \mu x A(x) \vdash \neg A(0)$. This is obviously not valid on the specified semantics.

What is required, then, is an independent argument as to why M fails. In Chapter 5, Field argues that if M held we could prove something obviously not true. Take a sorites sequence from the times when someone is young, say, to the times when they are old. Since there are times at which they are old then, applying M, we could establish that there is a first time at which they are old, which there is not. Now, it is certainly counter-intuitive to suppose that there is a first time of oldness – that, after all, is what drives the sorites paradox; but the 'forced-march' version of the paradox shows us that we are stuck with this (or something like) it anyway. As we backtrack down the series, asking a semantically competent respondent whether the person is old, their response must eventually change, and there must be a first point where this is so.[17]

In any case, the leastness of the object denoted by the description is, in fact, irrelevant. The paradox arguments can be run just as well with an indefinite description operator, ε. $\varepsilon x A(x)$ denotes any one of the things which makes $A(x)$ true (if there is such a thing), picked out, let us suppose, by the Axiom of Choice. This validates the ε-form of M, does not give rise to the LEM (even in the context of Intuitionist logic), and has no implications for sorites sequences.[18]

15 H. Field, 'Is the Liar Sentence Both True and False', 23–40 of JC Beall and B. Armour-Garb (eds.), *Deflationism and Paradox* (Oxford: Oxford University Press, 2005).

16 G. Priest, 'Spiking the Field Artillery', 41–52 of JC Beall and B. Armour-Garb (eds.), *Deflationism and Paradox* (Oxford: Oxford University Press, 2005).

17 G. Priest, 'A Site for Sorites', 9–23 of JC Beall (ed.), *Liars and Heaps: New Essays on Paradox* (Oxford: Oxford University Press, 2003).

18 See Priest, 'Spiking the Field Artillery'.

7. Rational Acceptance

A problem for those who endorse a consistent account of paradoxes according to which some sentences are LEM-violating, 'indeterminate', 'fuzzy', 'defective' (however one wishes to express it), is how exactly to characterise this class of sentences, in order to say something about them (such as that there are such things, or that some particular sentence is one of them) as, for example, Field does in the following: 'Our inability to prove certain of our rules unrestrictedly truth preserving is thus not due to ignorance about whether they unrestrictedly preserve truth, it is due to a 'fuzzyness' in the question....' (290).[19] One cannot say literally that the LEM fails for A if, as for Field, we have De Morgan laws; for $\neg(A \vee \neg A)$ entails $\neg A \wedge \neg\neg A$. Nor, given the transparency of truth, can one say that $\neg T\langle A\rangle \wedge \neg T\langle \neg A\rangle$, for this is equivalent.[20]

One way in which one might try to characterise such sentences is to deploy the notion of acceptance and rejection developed by Field (in 3.7). Acceptance is having a (subjective) probability above a certain threshold; rejection is having a probability below the co-threshold. Hence, a gappy sentence might be one such that a rational person neither accepts nor rejects it. (Clearly, what an arbitrary individual does with their subjective probability distributions is neither here not there.)

The machinery of acceptance and rejection produces its own contradictions, however. Consider a sentence, R, of the form 'It is not rational to accept R'. Suppose that one accepts it; then one accepts R and 'it is not rational to accept R'. This, presumably, is irrational. Hence it is not rational to accept R. That is, we have just proved R. So it is rational to accept it.[21]

Let me spell out the argument in more detail. Let us use $Rat(A)$ to mean that it is rationally permissible to accept that A. There is one premise, namely, that, for all A, $\neg Rat(A \wedge \neg Rat(A))$. Let us call this premise P. P appears very plausible: if someone believes A, and, at the same time, believes that it is not rationally permissible

[19] See also the discussion, (75) of the limits of iteration of the determinacy predicate.

[20] For similar reasons, Field cannot assert things which one might have expected, such as the non-existence of dialetheias, since $\neg\exists x(Tx \wedge \neg Tx)$ is equivalent to $\forall x(Tx \vee \neg Tx)$. Field has instead to deny $\exists x(Tx \wedge \neg Tx)$. More on denial in due course.

[21] This is a version of the 'irrationalist's paradox'. See, further, G. Priest, *Doubt Truth to be a Liar* (Oxford: Oxford University Press, 2006), 6.6.

to believe A, that would seem to be pretty irrational – not something that is itself rationally permissible.[22] Now, let $R = \neg Rat(R)$. We have the following deduction:

$$\forall A \neg Rat(A \wedge \neg Rat(A))$$
$$\neg Rat(R \wedge \neg Rat(R))$$
$$\neg Rat(R \wedge R)$$
$$\neg Rat(R)$$
$$R$$

There are a few steps in the deduction about which one might cavil generally, but nothing that seems supportable in this particular case. So we have $P \vdash R$. Assuming that deducibility is closed under *Rat*, we have $Rat(P) \vdash Rat(R)$. Again, one might have some worries about rational belief being closed under single-premise deducibility, but it is hard to apply them to this particular case. Now, it would certainly appear that it is rationally permissible to accept that P: just consider the case I gave you for it. Hence, we can infer $Rat(R)$, and we have a contradiction. A dialetheist can accept this; but not, of course, Field.

On 77–78 Field considers a paradox closely related to the one I have just given. His sympathy there is to question LEM for the paradoxical sentence. A virtue of the above articulation of the argument is that it makes it clear that it nowhere appeals to $R \vee \neg R$. Nor does it even help to contest $P \vee \neg P$. The only, somewhat desperate, move is to reject P.

8. Determinacy

In any case, Field has no desire to express fuzzyness in terms of rational acceptance (78), so he takes a different route. His strategy involves defining an operator, DA, to be read 'A is determinately true'. This is defined as $A \wedge (\top \rightarrow A)$ (where $\top$ is a logical constant that always takes the value 1); DA is true in a model iff A has ultimate value 1 in the model. The properties of determinacy include (236):[23]

1D $\models DA \rightarrow A$
2D $A \models DA$
3D If $\models A \rightarrow \neg A$ then $\models \neg DA$

[22] One might also interpret $Rat(A)$ as 'it is rationally obligatory to accept A'. The premise then has less plausibility. It would certainly seem quite coherent to believe A, but that it is not rationally obligatory to accept A. And one might argue that it is rationally obligatory to believe *that*.

[23] It is worth noting that the validity of 2D requires the inference $A \vdash B \rightarrow A$. This is one thing that drives Field down the irrelevant logic path.

It is now easy to check that in the semantics, the liar sentence, L_0, is not determinately true or false: $\vDash \neg DL_0 \wedge \neg D\neg L_0$. Bringing in determinacy always raises worries of extended paradox. What of the sentence, L_1, of the form $\neg DT\langle L_1\rangle$, 'This sentence is not determinately true'? We cannot have $\neg DL_1$, on pain of contradiction. But a little computation shows that $\vDash \neg DDL_1 \wedge \neg DD\neg L_1$. And so on. Let us write D^n for n consecutive 'D's ($n > 0$). Then for any finite n, there is a liar-like sentence, L_n, of the form $\neg D^n T\langle L_n\rangle$. We cannot have $\neg D^n L_n$, but we do have $\vDash \neg D^{n+1} L_n \wedge \neg D^{n+1} \neg L_n$.

But can we not iterate the process into the transfinite?—and do not then problems arise when we construct a sentence that says of itself that it satisfies none of these? The book contains a long and technical discussion of the matter. The answer is that we can indeed iterate the construction into the transfinite in a certain way, but that this does not generate novel contradictions; the reason is essentially as follows. Suppose that we had a hyper-determinacy operator, H, satisfying the following conditions:

1H $\vDash HA \rightarrow A$
2H $A \vDash HA$
3H $\vDash HA \rightarrow DHA$

Then inconsistency would ensue, without any appeal to LEM, for the following reasons. Let $Q = \neg HT\langle Q\rangle$. Then $\vDash HQ \rightarrow Q$ (by 1H), i.e., $\vDash HQ \rightarrow \neg HQ$. So $\vDash \neg DHQ$ (by 3D). Hence, $\vDash \neg HQ$ (by 3H), i.e., $\vDash Q$. So $\vDash HQ$ (by 2H). Now, it might be thought that if we had the resources of an infinitary conjunction with ω conjuncts, we could define such a hyper-determinacy operator. Consider the infinite conjunction CA: $DA \wedge D^2A \wedge D^3A \wedge \ldots$ Assuming standard properties for conjunction, we appear to have a hyper-determinacy operator since:

1. $\vDash CA \rightarrow DA \rightarrow A$
2. $A \vDash DA \vDash D^2A \ldots$ So $A \vDash CA$
3. $\vDash CA = (DA \wedge D^2A \wedge D^3A \wedge \ldots) \rightarrow (D^2A \wedge D^3A \wedge \ldots) \rightarrow$
   $D(DA \wedge D^2A \wedge \ldots) = DCA$

The argument for the last of these fails, since $\nvDash (DB \wedge DC) \rightarrow D(B \wedge C)$. (Though we do have $(DB \wedge DC) \vDash D(B \wedge C)$.) However far we extend into the transfinite, matters will be essentially the same.

So far so good. But does this settle matters? There are two jobs for the notion of defectiveness to do: [a] we must be able to say of certain sentences, e.g., the liar, that they are of this kind; and [b] we must be able to talk about such sentences in general and say things about them. The reason for [b] is well explained by Field himself. After discussing the possibility of using acceptance and rejection to express the notion of defectiveness in *KFS* (the theory of a Kripke fixed point with the logic K_3), he says (76ff):

> I think that this does *not* fully avoid the problem: I think that we need a notion of determinate truth, or equivalently, of a weakened negation (or of defectiveness). Rejection is inadequate precisely because it doesn't correspond to an embeddable operator: confining ourselves to it would cripple what we can say...
>
> For instance, consider cases where the notion of defectiveness appears in the antecedent of a conditional... ('If some of the sentences asserted in the chapter are defective, some reviewers are bound to point this out.') Or cases where the notion of defectiveness appears in a more highly embedded manner. ('There are theories of truth that don't contain defective sentences that are better than all theories that do contain defective sentences.') Debates about what is defective and what isn't would be hard to conduct without embedded defectiveness claims...
>
> ... the inability of *KFS* to express the notion of determinateness is a crippling limitation.

Though the conclusion concerns *KFS*, the reasons given are, note, quite general.

Now, Field's D operator does the job of [a] in many cases. As we have seen, for a number of defective sentences, A, like the Liar, we can say truly $\neg DA \wedge \neg D \neg A$, or at least $\neg D^{\alpha} A \wedge \neg D^{\alpha} \neg A$, for some suitable iteration of 'D's – maybe into the transfinite. But the D operator cannot do the job of [b]: as α increases, the extension of $\neg D^{\alpha}$ gets larger and larger (p. 238), so for no α does the extension of $\neg D^{\alpha}$ comprise all the non-(determinately true) sentences. Where $Q = \neg D^{\alpha} T\langle Q\rangle$, Q is not in the extension of $\neg D^{\alpha}$. Nor is it possible to define a predicate whose intuitive meaning is something like $\bigvee\{\neg D^{\alpha} Tx$: α is an ordinal$\}$, since, as Field shows, the precise definition of this depends on some ordinal notation, and will therefore take us only so far up the ordinals.

At this point, a different strategy suggests itself. The notion of having ultimate value 1 (with respect to a ground model) is expressible in the language. Does this not express the notion of being determinately true? No: its extension cannot line up with the really determinately true sentences, as Field is often at pains to stress. For suppose it did. Consider a sentence B, of the form $|B| \neq 1$. Suppose that $|B| = 1$; then B is determinately true, and so true: $|B| \neq 1$. Since we have LEM for this notion, we have established that $|B| \neq 1$. Hence we have established B. So B is determinately true: $|B| = 1$. Contradiction.

It would seem, then, that we need a new operator, Δ, to express the general notion of determinate truth. (So that A is indeterminate if $\neg\Delta A\wedge\neg\Delta\neg A$.) This threatens a revenge paradox. Consider a sentence, F, of the form $\neg\Delta T\langle F\rangle$. We would expect that $\Delta F\vdash F$ ($=\neg\Delta F$). Hence we have established $\neg\Delta F$ (by LEM), i.e., F. So ΔF. Contradiction. One might deny the application of LEM here: F is not determinate, and so we cannot apply the LEM in this instance. But to say that F is not determinate is to say that $\neg\Delta F\wedge \neg\Delta\neg F$, and this cannot be endorsed: for it entails $\neg\Delta F$, and so $\Delta F\vee\neg\Delta F$. So the status of F is inexpressible. We have therefore satisfied [b], but now [a] has gone.

In other words, there is no way that there can be a defectiveness predicate which will do the jobs required of it consistently. We can have [a] or [b], but not both. Or, to put it in a more familiar form, Field can maintain consistency only at the cost of expressive incompleteness.[24]

Field's response to this whole situation is simply to reject Δ (22.4–22.6), taking the notion to be 'ultimately unintelligible' (356). He does recognise this as a 'genuine cost', but one that can be paid (23.2). It cannot; for exactly the reasons that he himself gave in connection with KFS. Indeed, without the notion, one cannot even formulate the driving thought behind Field's own solution: that the LEM fails because of the existence of indeterminacies.[25] To declare all general claims about indeterminacy unintelligible is an act of ladder-kicking-away desperation of *Tractarian* proportions.[26]

[24] And just to prove that Field hasn't got all the good musical references to vendetta:

Revenge, oh, sweet revenge
is a pleasure reserved for the wise;
to forgo shame, bold outrage,
is base and utter meanness.
With astuteness, with cleverness,
with discretion, with judgment
if possible. The matter is serious;
but believe me, it shall be done.

Marriage of Figaro, Act 1.

[25] And as we see in a moment, the model theory, on Field's conception, cannot be invoked to explain the failure of LEM either.

[26] See, further, G. Priest, 'Spiking the Field Artillery'.

9. The Role of Model Theory

A naive assumption would be that Field's definition of model-theoretic validity is a definition of validity *tout court*. However, this would be a mistake. Field is clear that the model-theoretic notion is distinct from the real notion. (See, e.g. 67.) What, then, is real validity? Field does not address the question at any length in the book. His views on the matter can be found in a later paper.[27] The idea here is not to define (real) validity, but to take it as a primitive notion, and to axiomatize a theory about it, the norms of belief, and other related notions. Thus (to give a simple example) we might have an axiom such as: If one knows that the inference from A to B is valid, then one's degrees of belief, δ, ought to be such that $\delta(A) \leq \delta(B)$.

Let us grant, at least for the sake of argument, that this is the real notion of validity. To solve the paradoxes, we need to know which arguments are really valid, and which are not. Why should we take the results of the model theory to tell us anything about real validity? It would appear, at the very least, that there is a lacuna in Field's case here.

Field (he tells me) sees the matter in the following way. The considerations marshalled in the book, especially in its first two parts, are to be thought of as supporting the conclusion that the real notion of validity validates the T-schema, not LEM, etc. The model-theoretic construction is then but a *model* of the real notion (in the scientist's sense, not the logician's). Fair enough. But it is important to note that this approach gives up taking model theory in the way in which it is usually thought of. Standardly, one thinks of the truth (-in-an-interpretation) conditions of connectives as spelling out (an aspect of) its meaning. The meaning-conditions explain why an inference involving the connective is valid, and thereby justify it. In Field's approach, the model theory plays no justificatory role of this kind; it is, as Dummett puts it, a 'merely algebraic' semantics.[28] This is, of course, a coherent way to look at (logician's) models, but it deprives model theory of the ability to answer various questions about meaning, justification, etc. Of course, one may attempt to answer these in another way; but this still needs to be done. And in

[27] H. Field, 'What is the Normative Role of Logic?', *Proceedings of the Aristotelian Society, Supplementary Volume* **83**:251–68.

[28] M. Dummett, 'The Justification of Deduction', *Proceedings of the British Academy* **59** (1975), 201–32; reprinted as ch. 17 of *Truth and Other Enigmas* (London: Duckworth, 1978). See, further, G. Priest, 'Is the Ternay R Depraved?', a paper given at the conference *Foundations of Logical Consequence*, St Andrews, 2009.

evaluating rival theories of some notion, it is necessary to remember that one needs to look at the big picture in which it is embedded. It is no good, for example, to have a simple and elegant theory of truth, if the cost is a wildly implausible theory of meaning, or no acceptable account of *why* a valid inference, given a belief in the premises, reasonably grounds belief in the conclusion. I am not suggesting that Field's theory is of this kind, but simply that we have, as yet, a very partial theory, which needs a lot more to be said about it.

10. Model theoretic validity and (real) validity

But in any case, how good a model of real validity is Field's model-theoretic definition? One might hope that it is at least extensionally equivalent, but it is not clear that this is so. The problem is that we often reason about situations which concern all ordinals. Indeed, reasoning in *ZF*, and so Field's own model-theoretic reasoning, is of this kind. However, there is no model which contains all ordinals, so this situation is not within the compass of the model-theoretic sweep. Moreover, it is not just that there *is* no 'intended interpretation': there *cannot* be. There can be no interpretation in which taking the designated value coincides with truth. If there were an interpretation where, for all A, $|A| = 1$ iff A, we would have a contradiction, since set-theoretic statements such as $|A| = 1$ are bivalent. Field himself frequently stresses this point, e.g. (67–68):

> I wish I had something helpful to say about what makes a construction like Kripke's illuminating, but I don't. I hope I've already made clear, though, that one thing one *can't* easily say is that it is illuminating because it involves the notion of truth. For it involves instead a technical notion of having semantic value 1 (relative to a model), and this cannot in general be identified with truth without commitment to an extraordinarily contentious doctrine.[29]

Hence, the model-theoretic notion seems explicitly *barred* from applying to global reasoning. Plain *ZF* is beset with exactly the same problem. According to *ZF*, there is no universal set, and hence no standard model, and so no model for the model-theory to take in its scope.[30]

[29] GP: The doctrine is the indefinite extensibility of set-like entities. See the last paragraph of p. 64.

[30] For more on this matter, and its connection with 'revenge paradoxes', see G. Priest, 'Revenge, Field, and *ZF*', ch. 9 of JC Beall (ed.), *Revenge*

Field is aware of the matter. He observes that in the case of *ZF*, one may attempt to solve this problem using Kreisel's 'squeeze argument'.[31] Given an axiomatic proof-theory for the logic, we argue (by checking the axioms and rules) that if something is provable, it is intuitively valid, so it applies to everything, and so to all set-theoretic models, and so is provable (by the completeness theorem for the logic). Since we have now gone round in a loop, these notions are all extensionally equivalent. Field notes, however, that the corresponding strategy will not work in his case, since the model-theoretic notion of validity is not axiomatisable.[32]

Field's response to the situation is optimism (355):

> The fact that standard model theories of *ZF* and [my logic] don't allow the real world to be a model is, of course, something that I too have repeatedly emphasized . . .: it is why we have no intuitive *guarantee* that the model theoretic explications of validity are extensionally correct, though we may have a reasonable conviction that they are. (The fact that the real world is excluded from being a model could only affect the validity of sentences with unrestricted quantifiers, so the impact of a possible extensional failure is limited. Still, it's there.) A guarantee against extensional failure would be nice, but it's a mean old world. . .

It is indeed. The model-theoretic notion and the real notion are not extensionally equivalent. Suppose they were. Then an inference is not (really) valid iff it is not model-theoretically valid. So it is either (really) valid or it is not, since the model-theoretic notion is bivalent. But the real notion cannot be bivalent for Field. To see why, consider the following argument, Φ:

$$\frac{\top}{\Phi \text{ is (really) invalid}}$$

of the Liar: New Essays on the Paradox (Oxford: Oxford University Press, 2009).

[31] G. Kreisel, 'Informal Rigour and Completeness Proofs', 138–171 of I. Lakatos (ed.), *Problems in the Philosophy of Mathematics* (Amsterdam: North Holland Publishing Company, 1967).

[32] The matter is not entirely straightforward. Since Field takes the ground model to contain the standard model of arithmetic, of course it is not axiomatizable. The point (I take it) is that the notion of validity is not recursively enumerable with respect to an oracle that delivers the sentences true in the standard model.

We may reason as follows. The argument uses nothing but classically valid inferences plus two inferences concerning the predicate 'valid':

$$\frac{\langle A, B\rangle \text{ is valid} \quad A}{B} \qquad \frac{\begin{array}{c}\overline{A}^{(1)}\\ \vdots \\ B\end{array}}{\langle A, B\rangle \text{ is valid (1)}}$$

where, in the second of these, B depends on no undischarged assumptions other than A, the overlining indicates that A is discharged, and the number indicates the line at which A is discharged.

$$\dfrac{\dfrac{\dfrac{\overline{\Phi \text{ is valid}}^{(1)} \quad \dfrac{\overline{\Phi \text{ is valid}}^{(1)} \quad \overline{\top}^{(2)}}{\Phi \text{ is invalid}}}{\bot}}{\Phi \text{ is invalid (1)}}}{\Phi \text{ is valid (2)}}$$

But:

$$\frac{\Phi \text{ is valid} \quad \top}{\Phi \text{ is invalid}}$$

Hence, $\top \vdash (\Phi \text{ is valid} \wedge \Phi \text{ is invalid})$, and so, classically, $\vdash \bot$. Now, let V be: Φ is valid $\vee$ Φ is not valid. Then, for Field, $V \vDash \bot$. (We of course have $\top \vee \neg \top$.) Hence, Field must reject V.[33]

Field considers a version of this argument (p. 305 ff.), and makes, in effect, two objections. One is to query the first inference involving the predicate 'valid'. This is a hard move to sustain. If the first premise of the inference is true, the conclusion follows from the second premise alone. So if the premises are true, the conclusion must be. The second objection is to line (2). We have deduced (1) from $\top$ *plus* V. And V, though true, is not a logical truth; so we cannot infer (2). This, I think, misses the point. Given the legitimacy of the inference for the predicate 'valid', all the reasoning, including the reasoning establishing (2), is classically correct. This is all one needs.

[33] Though he takes this to be a bad thing: 'It would seem to be somewhat detrimental to the role of logic as a regulator of reasoning if we were unable to say that any given piece of reasoning is either valid or not valid' (307).

Despite his expression of hope above, Field, in fact, seems to be ready to concede that the extensional equivalence between real validity and model-theoretic validity may fail, at least in one direction. He says that since the model-theoretic notion is not axiomatisable, it might be preferable to take the really valid inferences to be a subset of the model-theoretically valid inferences. Thus (277):

> One issue I haven't discussed is *how many* of the inferences that preserve value 1 really ought to be declared logically valid *tout court* (as opposed to validated by the formal semantics). If one says that they *all* should, then one will need to make a decision on some seemingly arbitrary features of the semantics, such as the choice of the starting valuation for conditionals ... in order to decide what is "logically valid". In addition, the set of "logically valid inferences" will have an extremely high degree of non-computability. It might be better to adopt the view that what is validated by a given version of the formal semantics outruns "real validity"; that the genuine validities are some effectively generable subset of those inferences that preserve value 1 in the given semantics.

(Note that this does not resuscitate the squeeze argument, since the logic is now not complete). So the reader is left puzzling not only about how to do many of the things that model theory is usually taken to do, but also about which inferences really are valid.

11. The *T*-Schema

There is a further problem for Field concerning conditionality, which I will come to in due course; but I now want to turn from a discussion of Field's own view to his criticisms of a dialetheic account of the paradoxes. Let us take these chapter by chapter.[34]

Chapter 25 concerns the *T*-schema, $T\langle A\rangle \leftrightarrow A$. Field and I agree in endorsing this. Unlike Field, however, I reject the contraposed *T*-schema, so I do not endorse $\neg T\langle A\rangle \leftrightarrow \neg A$ in general.[35] Field objects: one of the main arguments for the *T*-schema is that it gives

[34] In what follows, I will refer to my *In Contradiction* (Dordrecht: Martinus Nijhoff, 1987; second edn., Oxford: Oxford University Press, 2006) as IC, and my *Doubt Truth to be a Liar* (Oxford: Oxford University Press, 2006) as DTBL.

[35] Though I do endorse it in one direction: $\neg T\langle A\rangle \rightarrow \neg A$ (IC, 4.7). And a dialetheist about the paradoxes certainly *can* accept it in both, as does JC Beall, *Spandrels of Truth* (Oxford: Oxford University Press, 2009).

us a way of expressing what someone said whilst unable to remember or, for other reasons, articulate what this was. Thus, suppose we say 'If everything that the Conyer's report says is true, then *B*'. Given that the Conyer's report says $A_1, \ldots, A_n$, 'this had better be equivalent to "If A_1 and ... and A_n, then *B*"' (210). Similarly (363), if we say 'if it's not the case that everything in the Conyer's report is true, then *B*', that is:

(1) $(\neg T\langle A_1\rangle \vee \ldots \vee \neg T\langle A_n\rangle) \rightarrow B$

this should be equivalent to 'If $\neg A_1 \vee \ldots \vee \neg A_n$ then *B*'. So, in particular, we can infer *B* from, say, $\neg A_1$.

Now, I entirely agree that an important function of the *T*-schema is as a device which 'strips off quotes', and so allows us to express things we cannot articulate (IC, 4.2). However, note that the effect that Field indicates can be obtained without the contraposed *T*-schema. Instead of saying 'if it's not the case that everything in the Conyer's report is true, then *B*' we can simply say 'if something in the Conyer's report is false, then *B*'. That is:

(2) $(T\langle \neg A_1\rangle \vee \ldots \vee T\langle \neg A_n\rangle) \rightarrow B$

Given just the ordinary *T*-schema, this is equivalent to 'if $\neg A_1 \vee \ldots \vee \neg A_n$ then *B*', and we can infer *B* from $\neg A_1$. Moreover, consider the claim that falsity entails untruth:

$T\langle \neg A\rangle \rightarrow \neg T\langle A\rangle$

If one rejects this (as I do), then, given only (1), one should not even want to move from $\neg A_1$ to *B*: one may have $\neg A_1$ whilst rejecting $\neg T\langle A_1\rangle$. To say something which legitimises the move for $\neg A_1$ to *B*, one needs precisely (2).

Note that if one has the full contraposed *T*-schema, then gluts are also gaps, in the sense that $T\langle A\rangle \wedge F\langle A\rangle$ $(=T\langle \neg A\rangle)$ entails $\neg T\langle A\rangle \wedge \neg F\langle A\rangle$. Field says that this is my reason for rejecting it (364–5). It isn't. The reason is that accepting it would turn any contradiction, $A \wedge \neg A$, into one of the form $T\langle A\rangle \wedge \neg T\langle A\rangle$, and 'contradictions should not be multiplied beyond necessity' (IC, 4.9). In general, if *A* is true and false, I see no reason to suppose that $T\langle A\rangle$ should be false as well as true (IC, 5.3).

On a more substantial point, I have given a non-triviality proof for the *T*-schema (and self-reference) based on a suitable relevant logic.[36] Field says (p. 371) that this is 'uninteresting' since in the model the *T*-schema contraposes. This is a strange claim. The non-triviality

[36] Priest, 'Paraconsistent Logic', 8.1, 8.2.

proof was supposed to do just that: prove non-triviality; and if the theory with the contraposed T-schema is non-trivial, so is the one without it, since this is weaker. However, in the semantics I have given for the truth predicate (IC, 5.4), the truth value of A does not determine the truth value of $T\langle A\rangle$. (For all A which are both true and false, $T\langle A\rangle$ is true; for some it is false as well.) It is therefore a fair enough question to ask (as Field does, 372) what does.

There are a number of ways in which one might answer the question, but one natural way is as follows. We take our cue from the thought that something should not be taken to be contradictory unless we are forced to suppose so. Now, though we are forced to take some sentences of the form $T\langle A\rangle$ to be contradictory when A itself contains 'T' (e.g. when $A = \neg T\langle A\rangle$), nothing seems to require us to take them to be so when it does not. Hence we may suppose that when A is true and false, $T\langle A\rangle$ is true, and if A contains 'T' (but not otherwise) it is false as well. It is easy enough to modify the fixed-point construction involved in the non-triviality proof to embody this idea.[37]

12. Conditionals

Field also objects to the conditional I use in formulating the T-schema. According to me, some instances of the T-schema are false (as well as true), since $A \wedge \neg B \models \neg(A \rightarrow B)$. Field says – without argument – 'I'm inclined to the view that the only satisfactory way of validating a truth schema within a dialetheic logic is to reject that the instances of the T-schema are dialetheias' (369; see also 373). Well, I'm not.

Field then goes on to say (373):

> [M]y main reason for being unhappy about Priest's dialetheic theories [of the conditional] ... is that conditionals of the form $A \rightarrow B$ behave very oddly *even when A and B are in the ground language* (the language not containing 'True', 'instantiates', etc.)
>
> Part of the point here is that these conditionals do not reduce to the classical conditional when the antecedent and the consequent are in the ground language...
>
> But a more important part of the point about the odd behaviour of Priest's conditionals is that they behave very differently

[37] Essentially, at each successor stage, we put A in the extension of 'T' if it is true at the previous stage; and we put A in the anti-extension of 'T' if it is false at the previous stage *and* contains 'T'. Otherwise, details remain the same.

> *even from the classical modalized conditional* ($A \rightarrow B$), and *even in the ground language.*

Indeed so, and this is all to the good. The material and strict conditionals are nothing like a real conditional, which is a relevant conditional.[38]

In truth, as far as the conditional goes, the boot is on the other foot. A feature of Field's logic, as we have noted, is that his conditional collapses into the material conditional in the ground language. But this means that it is open to the standard counter-examples given by relevant logicians.[39] Field recognises that there is a problem here. He briefly states his view concerning the 'paradoxes of material implication' as follows (374):

> embedded conditionals in the ground language behave in accordance with classical logic; but... in a typical utterance of an unembedded conditional "If A then B", the 'if... then' isn't really an operator at all. Instead, the unembedded conditional is to be evaluated according to the Ramsey test: the assertion is legitimate if and only if the conditional probability of B given A is high.

The problems with the material conditional go a long way beyond the simplistic $A \vDash B \supset A$ and $\neg A \vDash A \supset B$, however; they involve nested conditionals. To give one very standard example: suppose that we have a simple electrical circuit, in which a light, *c*, is in series with a battery and two switches, *a* and *b*. The light is on iff both switches are closed. Let A be '*a* is closed', B be '*b* is closed', and C be '*c* is on'. Reading '$\supset$' as 'if', we have $(A \wedge B) \supset C$. Classical logic assures us that the obviously false $((A \wedge \neg B) \supset C) \vee ((\neg A \wedge B) \supset C)$ follows. (And note that this has nothing to do with vagueness either.)

Nor can Field claim it as an advantage of his approach, *vis à vis* a relevant conditional, that it allows the recapture of 'classical' reasoning in a certain way. If we are reasoning about a consistent and complete situation, the following principles of inference, though not valid, are materially correct: $A \vdash B \vee \neg B$, $A \wedge \neg A \vdash B$. The addition

[38] Field also goes on to point out that the relevant conditional is nothing like the Lewis/Stalnaker conditional, since it satisfies strengthening of the antecedent ($A \rightarrow C \vdash (A \wedge B) \rightarrow C$), etc. The Lewis/Slalnaker conditional is a *ceteris paribus* conditional; and there are natural relevant versions of these, just as much as there are strict *ceteris paribus* conditionals. See Priest, *Introduction to Non-Classical Logic*, 10.7.

[39] See Priest, *Introduction to Non-Classical Logic*, 1.9; and for a comprehensive list, R. Routley, V. Plumwood, R. K. Meyer, and R. T. Brady, *Relevant Logics and their Rivals* (Atascadero, CA: Ridgeview, 1982), ch. 1.

of these to a relevant logic ensures that the $\rightarrow$-free fragment is classical logic.[40]

I might also add that the semantics of relevant logic are much, much, simpler than those of Field. Critics of relevant logic have often averred that they are too complex to be taken seriously. I think that the semantics for the basic systems of relevant logic are, in fact, not particularly complex: they simply invoke a certain notion of impossible world, needed for other purposes anyway.[41] But Field's semantics are more complex by several orders of magnitude.

13. Soundness and truth preservation

Chapter 26 concerns issues to do with soundness and truth preservation. I think that arithmetic is inconsistent. The arguments are given in IC, Chapters 3 and 17. Field says, without discussion of these arguments, that the view 'defies belief' (377); he give no reasons as to why it should. Now, it is true that arithmetic has not been beset by paradox in the way that set theory has. But anyone contemplating the standard proof of Gödel's Theorem must have been struck by the thought that if, say, Peano Arithmetic is consistent, it is so only by 'good luck'. In all honesty, Gödel's paradox is just as problematic for arithmetic as Russell's paradox is for set theory.

As we have seen, the paradox concerns the sentence, G, 'This sentence is not provable'. If G were provable, it would be true, and so not provable. Hence it is not provable. But this is a proof of G. So it is provable. The argument seems compelling for the naive notion of provability. It depends on two principles concerning provability:

1. if $\vdash A$ then $\vdash Prov\langle A\rangle$
2. $\vdash \forall x(Prov\ x \rightarrow Tx)$

IC, Chapter 3, outlines a proof of 2 in a semantically closed arithmetic. The argument is by a standard induction over the length of proofs. As we have already noted, and Field points out (377), the natural way of arguing requires:

(*) $(T\langle A\rangle \wedge T\langle A \rightarrow B\rangle) \rightarrow T\langle B\rangle$

[40] In fact, my preferred way of recapturing classical reasoning is not this, but uses a non-monotonic paraconsistent logic which behaves classically on consistent sets of premises. See IC, ch. 16.

[41] See DTBL, p. 127 f., and for technical details, Priest, *Introduction to Non-Classical Logic*, chs. 9 and 10. For an account of the meaning of the much maligned ternary relation, see Priest, 'In the Ternay *R* Depraved?'.

to handle the case for MP; one cannot have this for Curry reasons. Now, axiom systems do not have to use MP, of course. Thus, the proof procedure for the finite inconsistent arithmetics is an algorithm which does not invoke an axiom system at all (see IC, 17.2). But even if 2 is not provable, it is still, presumably, true (in fact, analytic if 'prove' just means 'establish as true'). And so we can take it as an axiom if necessary. It is, then, possible for a dialetheist to have what, as we have noted, one would expect: a semantically closed arithmetic which endorses its own soundness (contradictions and all).[42]

Turning to a different matter concerning truth preservation: Neither Field nor I can endorse (*). Field points out (378) that since $\forall x \forall y((Tx \wedge T(x \rightarrow y)) \rightarrow Ty)$ entails triviality, it entails its own negation. By the LEM (which I accept), I must accept its negation. So I must accept that MP is not truth-preserving in this sense; and so I do. But truth preservation is said in many ways. For a start, even if $A \vDash B$ does not entail $T\langle A \rightarrow T\langle B\rangle$, it does entail that $T\langle A\rangle \vDash T\langle B\rangle$. So we always have truth preservation in this form.

More substantially, like Field, I accept that MP is truth-preserving in the form: $\forall x \forall y(T(x \rightarrow y) \rightarrow (Tx \rightarrow Ty))$. This may not be the form in which truth preservation is used in a standard soundness proof, but it still serves as a statement of the facts. It might be retorted that the inference of adjunction, $A, B \vdash A \wedge B$, is not truth preserving in this sense, since $\nvDash T\langle A\rangle \rightarrow (T\langle B\rangle \rightarrow T\langle A \wedge B\rangle)$. But we do have $\vDash (T\langle A\rangle \wedge T\langle B\rangle) \rightarrow T\langle A \wedge B\rangle$. And we can obtain an expression of truth-preservation for two-premise inferences, $A,B \vdash C$, by disjunction:

$$((T\langle A\rangle \wedge T\langle B\rangle) \rightarrow T\langle C\rangle) \vee (T\langle A\rangle \rightarrow (T\langle B\rangle \rightarrow T\langle C\rangle)).$$ [43]

The technique obviously generalises to inferences with more premises. But there is also an issue with fewer premises. As Field notes (382), I endorse the validity of the inference $0 = 0 \vdash C \rightarrow C$, but $T\langle 0 = 0\rangle \rightarrow T\langle C \rightarrow C\rangle$ (or equivalently, $0 = 0 \rightarrow (C \rightarrow C)$) is not valid, as it is for Field. The example is a special case of a more general phenomenon. $C \rightarrow C$ is a logical truth. In the semantics of relevant logic,[44] if $\vDash B$ then, for any A, $A \vDash B$; yet if A and B share

[42] If provability analytically entails truth, then 1 becomes contingent upon having the right system of arithmetical proof. But as the very argument concerning G shows, beyond some means of producing self-reference, the only non-logical axiom required is 2 itself. Contesting other axioms is therefore beside the point.

[43] Actually, the order of the premises is important here, so the disjunction had better have three disjuncts: $((T\langle A\rangle \wedge T\langle B\rangle) \rightarrow T\langle C\rangle) \vee (T\langle A\rangle \rightarrow (T\langle B\rangle \rightarrow T\langle C\rangle)) \vee (T\langle B\rangle \rightarrow (T\langle A \rightarrow T\langle C\rangle))$.

[44] As formulated in Priest, *Introduction to Non-Classical Logic*, ch. 10.

no propositional parameters $\nvDash A \rightarrow B$.[45] Relevant logicians have always been suspicious of a bald notion of logical truth. B is supposed to be a logical truth if it is a consequence of the empty set of premises. But nothing comes from nothing. So how can this be? A standard answer invokes the constant t, thought of as the conjunction of all logical truths (not to be confused with the conjunction of all actual truths, often represented by the letter t as well). Given this, we can define a conceptually happier notion of logical truth, $\vDash'$, as follows:

$\vDash' B$ iff $t \vDash B$

Clearly, given that $\vDash t$, the two notions of logical truth are extensionally equivalent: $\vDash' B$ iff $\vDash B$. And if $t \vDash B$ then, since B is a logical truth, $\vDash t \rightarrow B$, and so $\vDash' t \rightarrow B$. The definition obviously generalises. Concentrating on the one-premise case for simplicity, define:

$A \vDash' B$ iff $A \wedge t \vDash B$

As before, $A \vDash' B$ iff $A \vDash B$. Moreover, if B is a logical truth, since $\vDash t \rightarrow B$, then, whatever A is, $\vDash (A \wedge t) \rightarrow B$, and so $\vDash' (A \wedge t) \rightarrow B$. We do have a logically valid conditional corresponding to the inference from A to B.

Finally on this topic, Field, as we have seen, can no more than I endorse the thought that valid inferences, $A \vDash B$, preserve truth in the sense that $T\langle A\rangle \rightarrow T\langle B\rangle$ is true. What is his solution to the problem? It is to endorse a restricted notion of truth preservation (288): if $A \vDash B$, then if we endorse A, we endorse $T\langle A\rangle \rightarrow T\langle B\rangle$. For if we endorse A, then we endorse B. So we endorse $T\langle B\rangle$, and so $T\langle A\rangle \rightarrow T\langle B\rangle$, since $D \vDash C \rightarrow D$. Field argues that I cannot have restricted truth preservation in the same sense, since the final inference is not available to me. True. But there is another form of restricted truth preservation. This time, let t be the conjunction of all *actual* truths. Then if $A \vDash B$, and we endorse A, then we endorse B, and so $t \rightarrow B$. Hence, we have $(t \wedge T\langle A\rangle) \rightarrow T\langle B\rangle$, a restricted form of truth preservation just as good as Field's.[46]

[45] Early texts on relevant logic concerned themselves only with logical truth, not logical consequence as such. To the extent that they had such a notion, to say that B was a consequence of A was simply to say that $A \rightarrow B$ is a logical truth. From the fact that B, it certainly does not follow that $\vDash A \rightarrow B$.

[46] For what it is worth, I also have truth preservation in the form $(T\langle A\rangle \wedge T\langle A \rightarrow B\rangle) \supset T\langle B\rangle$, which Field does not, since this depends on LEM.

Graham Priest

14. Revenge and Dialetheism

In Chapter 27, Field compares his view and mine with respect to revenge problems, and argues that to the extent that his account is subject to revenge problems, so is mine. Field's discussion is predicated on the assumption that my working metatheory is pursued in *ZF* (384, 385, 391, 392). It is not: it is to be taken as couched in a paraconsistent set theory (IC, ch. 18, esp. 259). This makes many of his comments beside the point. For example, Field claims (391) that my model-theory contains 'monsters' – connectives which make perfectly good sense in the semantics, but which must nonetheless be declared illegitimate – such as Boolean negation. Now, for me, Boolean negation is no monster; it is a connective with perfectly legitimate truth/falsity conditions. It is just that those conditions do not entail that the connective satisfies the rules of classical negation, as they do for Field; the rules of classical negation are needed to show that it does (DTBL, ch. 5).

More contentiously, perhaps: many people have argued that a dialetheic solution to the paradoxes can be maintained only by expressive incompleteness, particularly with respect to the notion of being *false only*.[47] Field is much more sympathetic on these matters that many critics, and even offers the dialetheist a hierarchy of 'false only' predicates, which mirrors his hierarchy of determinately-true predicates (388 ff.) – though I have no inclination to go down this path. He does, however, argue that he and I are in much the same boat in these matters. It looks as though he needs an indeterminacy predicate of a certain kind to say what he needs to say; it look as though I need a 'false only' predicate of a certain kind to say what I need to say. So I cannot claim any advantage in this matter.

Now, there are certainly parallels between our positions here. For a start, neither of us has a general way of *asserting* something with a certain effect. As we have seen, Field has no way of asserting that something is not true, if this is meant to include things that are false and things that are indeterminate; I cannot assert that something is false-only if this is required to exclude things that are true as well. For both of us, though, there is a way of obtaining this effect with a different kind of speech act, namely *denial*: both of us can deny that A is true.[48] In *some* contexts there are things that can be asserted which

[47] See IC, 20.4 for references and discussion.

[48] And if you trawl through both of our papers, you may well find a speech act of denial which employs the worlds 'indeterminate' (for Field)

have the same force as a denial. Thus, both of us can assert $A \rightarrow \bot$, in the face of which one can maintain A only on pain of triviality, which would normally be taken as a denial. As Field points out, though, this cannot be used to reject in all cases. The Curry sentence, K, for example, is itself of the form $K \rightarrow \bot$, and so cannot be asserted by way of rejecting K. With all of this I agree. There is nothing that can be asserted which will, in general, have the same force as denial. The speech act is *sui generis*.[49]

So far, our lines are parallel. However, on further projection, they diverge. The crucial difference concerns what can be said in propositional contexts. Field has literally no way of expressing the notion of indeterminacy. As we have noted, for him the notion is literally meaningless. The dialetheist about the paradoxes does have a way of expressing that something is false only – in the very words 'false and not true'. It is just that these cannot be guaranteed to behave consistently. Field is not persuaded. He replies (386): 'the problem is ... that on these definitions, the notions don't behave in accordance with how they seemed intended to behave when the theory was being explained', that is, consistently. But who intended them to be so? Not I. Field brings to the discussion a preconception of his own, namely that the metatheory is consistent; this is not mine. He says, by way of illustration (387):

> We'd like to say things like (1): if the premise of a conditional is solely true and the consequent solely false then the whole conditional is solely false; but for this to "mean what we'd like it to mean", then 'solely true' and 'solely false' in it had better mean "what we'd like them to mean"...

Well, they mean exactly what they say. Clearly, Field would like to pack consistency into the meaning. It's a fact of dialetheic life – indeed, of life in general – that you can't, however much you'd like too. It's mean old world.[50]

---

or 'false only' (for me). The question, as with all speech acts, is less about the words used, then about what is communicated in the act.

[49] See DTBL, ch. 6. In 6.3, I point out that asserting $A \rightarrow \bot$ cannot function as a denial in the mouth of a trivialist; nothing can.

[50] See, further, IC, 20.4. JC Beall has recently stressed to me that the sentences $A \rightarrow \bot$ and $\top \rightarrow A$ have many of the properties that the friends of consistency might expect 'false only' and 'true only' to have.

15. Intended interpretations

This deals with Field's criticisms. But given Field's difficulties concerning the extensional adequacy of his model-theoretic notion of validity, it is worth noting that a dialetheic approach does not share the same problems. For a start, the standard semantics for quantified relevant logics (with identity) are axiomatisable.[51] This means that the Kreisel squeeze argument could be used if it were needed; but it is not. Using a set theory with the unrestricted abstraction schema, there is no problem about the existence of the 'intended' interpretation. Thus, working within set theory, we can define the notions of an interpretation, truth in an interpretation ($\Vdash$), and the set of things which hold in an interpretation (Th), as usual, and show that $\forall x\ x \Vdash Th(x)$. We can then define the standard model, M, using naive comprehension:

$$\langle x, y\rangle \in M \leftrightarrow (\forall u\ u \in x \wedge \forall u, v(\langle u, v\rangle \in y \leftrightarrow u \in v))$$

and infer that $M \Vdash Th(M)$. Validity can also be defined in the natural way: an inference from A to B is valid iff $\forall x(x \Vdash \langle A\rangle \rightarrow x \Vdash \langle B\rangle)$.[52]

One thing one cannot have, as we have seen, is for the validity of the an arbitrary inference, $A \vDash B$, to entail truth preservation *in the form* $T\langle A\rangle \rightarrow T\langle B\rangle$, where $\rightarrow$ is a detachable conditional. That gives rise to triviality via Curry paradox. This means that we cannot expect to have $M \Vdash \langle A\rangle \leftrightarrow A$, in the form of a detachable (bi)conditional. We can, however, have it in the form $M \Vdash \langle A\rangle \equiv A$. And this is all we probably should expect, anyway. After all, M is not the actual situation; it is not a situation at all, but a set-theoretic representation of one. We should therefore expect no more than a material equivalence: the two sides are both true together or both false together. (Of course, even though $M \Vdash \langle A\rangle \equiv A$ is true, it is false as well if one of the sides is true and the other is false.)[53]

[51] See Priest, *Introduction to Non-Classical Logic*, ch. 24.

[52] See IC, 19.12, and ch. 18. There is, however, an issue of what, exactly, the set-theory is in which this is done. For a discussion see IC, ch. 18.

[53] I note that in 'On Dialethism', *Australasian Journal of Philosophy* **74** (1996), 153–161, L. Goodship has argued that the biconditional of the T-schema should itself be interpreted materially. (This does not mean that the language cannot contain a detachable conditional; just that this conditional is not used for the Schema.) This has some plausibility. The T-schema can be thought of as expressing no more than the claim that its two sides have the same truth value (both true or both false). And this is exactly what the material biconditional expresses. One of the virtues of

16. Putting the Pieces Together

I think that the main reason Field dislikes my solution to the paradoxes has, in fact, nothing to do with any of the above points. It is simply that the theory is inconsistent. That it is so, I of course agree. But as an *objection*, the point would have more force if Field actually *argued* that this is indeed a problem, and did not – untypically for him – simply take orthodoxy for granted.[54]

Concerning his actual arguments against my approach to the paradoxes, these were as follows:

- It does not endorse the contraposed T-schema.
- No semantics has been given for the non-contraposing T-schema.
- Its conditional is strange, and does not capture natural reasoning.
- It cannot express the truth-preservingness of valid inferences in an appropriate form.
- It is subject to expressability problems similar to those of his own theory.

These objections have now been answered. In turn, I have objected to Field's theory *vis à vis* mine that:

- It cannot formulate a theory according to which it, itself, is sound.
- It faces paradoxes, such as the Knower and the Irrationalist's paradox.
- The definability paradoxes have not been solved.
- The account still faces revenge problems connected with determinate truth.

---

this approach is that it provides a very simple solution to the Curry paradox, since MP fails for $\supset$. (Contraction holds.) A natural objection to the proposal is that it renders the T-schema impotent, since we can never get from one side of it to the other. In particular, the truth predicate cannot be used to make blind endorsements. The objection may not be as telling as it appears, however, since material detachment is still a valid default inference (see IC 8.5, and ch. 16). (The liar paradox, etc., are still forthcoming unconditionally.)

[54] See G. Priest, 'What's so Bad about Contradictions?', ch. 1 of G. Priest, JC Beall, and B. Armour-Garb (eds.), *The Law of Non-Contradiction: New Philosophical Essays* (Oxford: Oxford University Press, 2006), and the debate in many of the papers in that volume.

- It deploys a conception of model-theory which deprives this of a number of the functions which it is usually taken to have. How these are to be discharged still needs to be explained.
- His model-theoretic notion of validity and his real notion of validity are not coextensive; and it is not clear inferences are really valid.
- His conditional is subject to standard relevant counter-examples (and is complex).

Though some of these criticisms may just require further bits of the jigsaw, most are clearly more substantial.[55] When comparing the two accounts, then, it seems that a dialetheic account is (still) the rationally preferable one.[56]

On the cover of Field's book, there is a Raphael painting of a knight (St. George) fighting a dragon. In the background is a maiden. We are to suppose, I take it, that the knight represents Field, the maiden truth, and the dragon the paradoxes. The maiden, I fear, is not as innocent as she looks; she is, in fact, in cahoots with the wingèd reptile. My money's still on the dragon.[57]

Universities of Melbourne and St Andrews,
and the Graduate Center, City University of New York

[55] An interesting question is to what extent the consistent paracomplete theory of R. Brady, *Universal Logic* (Stanford, CA: CSLI Publications, 2006), based on a relevant conditional, fares any better. The account is not subject to the last three objections, but it is subject to the others.

[56] This locates Field's account in the more general discussion of DTBL, 7.5.

[57] This essay is based on a series of seminars given at the University of St Andrews in December 2008. Thanks go to the participants of the seminar for their thoughtful comments and criticisms, and especially to Ole Hjortland, Stephen Read, Stewart Shapiro, Crispin Wright, and Elia Zardini. Thanks also go to JC Beall and Stephen Read for written comments on an earlier draft. Especial thanks go to Hartry Field himself for extended email discussions on a number of matters dealt with here – and for many other fun discussions over the years and bottles of wine.