

Field's logic of truth

Vann McGee

Published online: 4 November 2009
© Springer Science+Business Media B.V. 2009

Saving Truth from Paradox is a re-exciting development. The 70s and 80s were a time of excitement among people working on the semantic paradoxes. There were continual formal developments, with the constant hope that these results would yield deep insights. The enthusiasm wore off, however, as people became more cognizant of the disparity between what they had accomplished, impressive as it was, and what they had hoped to accomplish. They moved onto other problems that they hoped would prove more yielding. That, at least, was how it seemed to me, so I was delighted to see a dramatically new formal development that is likely to rekindle our enthusiasm.

Field didn't build from scratch, of course. Since Thales, no one has. A construction upon which he particularly relied was given by Saul Kripke (1975), who applied the methods of first-order positive inductive definitions¹ to produce "fixed points" for a language with truth-value gaps, evaluations in which a sentence² φ always receives the same semantic status—true, false,³ or unsettled—as the sentence $T(\ulcorner \varphi \urcorner)$. Truth-value gaps were handled by the strong 3-valued semantics of Kleene (1952b, §54), so that a disjunction counts as true iff one or both disjuncts are true and as false iff both disjuncts are false, and similarly for the other connectives. Because the semantics is compositional, the Kripke construction exhibits a rare property JC Beall (2009) calls "transparency," according to which φ and $T(\ulcorner \varphi \urcorner)$

¹ See Moschovakis (1974) or Barwise (1975).

² We shouldn't really speak of a sentence (that is, a sentence type) as being either true or false, of course, but rather of it being true or false in a language in a context. For our discussion here, however, contextual variation won't be a much of a factor.

³ We identify falsity with the truth of the negation.

V. McGee (✉)

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 77 Massachusetts Ave 32-D931, Cambridge, MA 02139-4307, USA
e-mail: vmcgee@mit.edu

can be substituted anywhere, without affecting the semantic status of the containing sentence.

Transparency is an elusive goal. Kripke presents a number of variations on his basic fixed-point construction, based upon different methods of adapting classical model theory to languages with truth-value gaps. One such evaluation scheme, which Field discusses in an illuminating way, is the “strong supervaluation” method.⁴ Starting with a classical model of the ground language, and a consistent set E of sentences, which is to serve as the *extension* of “T,” and a set A , which is to serve as the *antiextension*, where A includes all the nonsentences and where no disjunction of sentences in A is a consequence of E , the method counts a sentence as true iff it is classically true in every expansion of the ground model in which a maximal consistent set of sentences that includes E and is disjoint from A is assigned to “T,” and it counts a sentence false iff it is false in every such expansion. With this alternative scheme for evaluating the semantic values of sentences, we still get a fixed point, but the fixed point isn’t transparent, because the semantics isn’t compositional. More precisely, with respect to the simple classification that gives the semantic status of a sentence as either “true,” “false,” or “unsettled,” we don’t have a compositional semantics. A subtler Boolean-valued semantics takes the semantic value of a sentence to be the set of classical expansions in which the sentence is classically true. With respect to this more sophisticated semantics, we do have compositional semantics, but we no longer have a fixed point. A compositional semantics that provided a fixed point would give us transparency, which we can’t have, if we have classical mathematics and rudimentary arithmetic. Gödel’s self-reference construction gives us a liar sentence λ , for which we can prove $(\lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner))$. Transparency would give us $(\lambda \leftrightarrow \neg \lambda)$, which is classically inconsistent.

The Kripke–Kleene construction achieves transparency, but at a cost. The Kleene logic is so weak that, as Feferman puts it (1984, p. 264), “nothing like sustained ordinary reasoning can be carried on.” Moreover, the fixed point omits such evident and innocuous truths as “Every true sentence is true.” You may want to deny the principle that you ought not assert something unless you regard it as true. Tim Maudlin (2006) has followed this path, arguing vigorously that assertability ought to outstrip truth. If you do accept the principle, however, you are likely to regard the stock of truths supplied by the Kripke–Kleene fixed points as so very meager that the construction is a hopeless dead end.

It turns out you’d be mistaken. Field has shown how to supplement the Kleene semantics for the truth-functional logical operators by adding a new conditional whose semantics is artfully designed so that something very much like sustained ordinary reasoning can be carried on in the resulting logic. The new conditional can’t behave classically, of course. Being transparent and accepting $(\lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner))$, the new semantics will accept $(\lambda \leftrightarrow \neg \lambda)$, which it regards as consistent. Utilizing the new conditional, Field develops a semantic theory that displays a combination of transparency and logical strength that I wouldn’t have thought possible. Frankly, I was astonished.

⁴ Adapted from van Fraassen (1966).

Field offers the same deal proposed by the Kleene version of the Kripke construction, to trade classical logic for transparency. But he's sweetened the alternative logic, and thereby made the bargain much more attractive.

The philosophical literature offers a wide menu of options, in addition to adopting Field's proposal or maintaining classical logic. It also offers a variety of ways to uphold classical logic, in addition to the strong supervaluational variant of the Kripke construction. But let me focus on just these two. In discussing their rivalry, I want to engage an issue that Field discusses only glancingly in the book, but that I think is important. I hope you will bear with me.

The issue is whether the theory of truth ought to be regarded as a part of linguistics, which in turn is a subdiscipline of social psychology, or as part of logic. It is not a question that initially looks promising. Disciplinary boundaries are mushy in general, and the linguistics/logic border is no exception. Even so, I think it is a question worth pursuing, for it will clarify what a theory of truth is aiming at.

In asking whether the theory of truth is part of social psychology, I have in mind the thesis that truth values and truth conditions play a central theoretical role in understanding how human beings communicate by language. The data the theory is intended to explain are entirely commonplace. Taking an example from Field's first published paper (1972, p. 23), Field's assessment of the likelihood that there is a foot of snow in Alabama is drastically altered by a trustworthy friend's utterance of the sentence, "There is a foot of snow in Alabama." To understand this change, we need to acknowledge the truth conditions of the sentence the friend asserted. We also need to recognize that the truth conditions of the utterance are the same as the truth conditions of the beliefs that the friend expressed and Field acquired. The doctrine that the truth values and truth conditions of sentences of a public language play a central theoretical role in linguistics goes hand in hand with the doctrine that the truth values and truth conditions of beliefs and other mental attitudes play a central role in psychology.

The alternative to thinking of truth as a theoretical notion belonging to the social sciences is to think of it as a merely logical device for simulating infinitary logic by finitary means. When we say "Everything the Pope says is true," we are, in effect, asserting the conjunction of infinitely many sentences of the form, "If the Pope says that φ , then φ ."⁵

Both ways of thinking about truth found an eloquent spokesperson in Field himself, at different stages in his career. The early Field (1972) embraced a social-science theoretic conception of truth, and he also embraced physicalism, holding that the organization of human societies and human minds is ultimately determined by the arrangement of matter. A commitment to classical logic isn't built into physicalism. Hilary Putnam, when he was still a physicalist (1965), proposed that classical sentential calculus breaks down at the quantum level. The idea didn't work out, but the problem was with the physics, not the metaphysics. However, if we reject Putnam's overtures and accept the traditional view that the techniques of classical applied mathematics are an integral part of the scientific method, as it is practiced in the physical sciences, to say that those same techniques are not

⁵ See Quine (1986, Chap. 1), Leeds (1978), and Field (2001, *passim*).

available when we are developing the theory of truth would be, I should think, incompatible with Tarski's goal of developing a theory of truth that is in "harmony with the postulates of the unity of science and of physicalism" (1936, p. 406). How well Tarski succeeded in this goal is controversial. Field (1972) argues that his efforts fell short, because some of his techniques would have been deemed physicalistically unacceptable if they had been employed within the natural sciences. The methodological strictness Field demanded would surely be incompatible with adopting a dual standard, whereby fully developed theories within the natural sciences have to be consistent, as consistency is gauged by classical mathematics, but semantic theory does not.

My complaint about methodological laxity assumes that semantic terms are the only ones treated nonclassically. For the most part, this is the framework within which Field works, but he toys with the idea of relinquishing classical logic whenever vague terms are in use.⁶ Outside of pure mathematics, nearly every term in our language is vague, to some extent, so the worry arises that methods of classical mathematics will be forbidden to us every time we step outside the math department. Field tells us not to worry, that it's OK to use classical applied mathematics except in very special circumstances, in much the way it's permissible to use Euclidean geometry unless the region we are examining contains unusually intense gravitational fields (2008, p. 105f). I am doubtful. We happily employ Euclidean geometry because general relativity theory assures us that, in the circumstances in which we employ it, the answers it gives are approximately correct. Perhaps there are theorems forthcoming that develop a nonclassical applied mathematics and prove that its results are approximated by classical applied mathematics, but until such results become available, I continue to worry. Because it's not entirely clear which people count as residents of California at any particular time, the phrase "the population of California in 2040" does not denote a definite number. Nonetheless, demographers are able to use the methods of classical statistics in remarkably sophisticated ways to make predictions about what the population will look like. Will these advanced statistical techniques still be available to us if we accept the judgment that the methods of classical mathematics are, strictly speaking, inapplicable outside of pure mathematics? Perhaps, but offhand it's hard to see how even such simple judgments as "the total population = the Latino population + the non-Latino population" can be obtained.

Terms from the social sciences are always vague, yet the traditional methods of mathematics are enormously effective, there no less than in the natural sciences. If you are a physicalist, that is what you'd expect. What you wouldn't expect is that you would have to relinquish those methods in order that your practice conform to some metaphysical doctrine about vagueness. A principal motivation of physicalism is to overthrow the domination of science by philosophy; physics, not metaphysics, should rule the roost.

⁶ People working on vagueness sometimes use a determinately true/determinately false/neither trichotomy to mark the cases in which truth conditions are underdetermined by speaker usage. Field develops a notion he calls "determinate truth," but it is a notion internal to his logical conception of truth, and it has no particular connection with speaker usage.

I want to conclude that, if you think that the theory of the truth conditions for statements and beliefs is a part of social science, and you are inclined toward materialism, you will resist the idea that classical logic is inapplicable within semantics. I know it's presumptuous to say this, but I suspect that the young, materialistic Field would not be entirely satisfied with what his more mature self is up to.

Field's views about why we need a notion of truth have changed. On his current understanding, "true" is a logical word, whose use enables us, as Quine (1984, p. 11) puts it, to express generality "along certain oblique planes that we cannot sweep out by generalizing over objects." It works by undoing the effects of quotation marks, and it works smoothly until we try to apply it to sentences that contain the word "true" or other semantic terms. Then the standard use of "true" comes into conflict with the standard use of the other logical terms, a conflict so severe that it threatens complete collapse.

We have no good procedures for resolving conflicts among logical principles. Within the other sciences, we can hope to settle disputes by logical argumentation, but logical quarrels question the procedures by which we settle disputes, and we have no higher arbiter. The worries about physicalism that exercised me earlier are no longer in play once we adopt a logical conception of truth, because logic treats all disciplines alike, and because logic isn't interested in questions of the priority of matter over mind. There is a basic requirement of conservativeness. The rules for using a new logical operator shouldn't generate new, unsupported claims that don't involve the operator. Once that requirement is met, the resolution of controversy is likely to consist in seeing which combination of rules works out best in practice. A tolerant attitude and an experimental spirit are called for.

In logical controversies, direct argument hasn't a large role to play, but it does have a role. You can prove by classical mathematics that classical first-order logic is sound. The simplest version of this theorem tells us that, if we have a first-order language \mathcal{L} for which we have a Tarski (1935)-style compositional theory of truth, we can prove in the theory of truth, together with a moderate portion of standard mathematics, that every sentence of \mathcal{L} that is a theorem of pure logic is true. But this method is only applicable if the sentences of \mathcal{L} don't contain the word "true" (assuming "true" here means "true in \mathcal{L} "); that's the fundamental limitation on Tarski's approach. So the theorem isn't helpful in answering the question whether classical logic can be faithfully applied in reasoning that involves the truth predicate. Without stepping outside the language \mathcal{L} , we won't be able to prove the general soundness theorem, but we will be able to prove the instances of a soundness theorem schema: ($\ulcorner \varphi \urcorner$ is provable in pure logic $\rightarrow \varphi$). For this, we need a theory, expressed in the language \mathcal{L} , that can describe the formation of finite sequences, that includes the Peano axioms (so we can do the Gödel coding), and that allows arbitrary formulas from \mathcal{L} to appear within instances of the induction axiom schema ($(\varphi(0) \wedge (\forall x)(\mathbb{N}(x) \wedge \varphi(x) \rightarrow \varphi(Sx))) \rightarrow (\forall x)(\mathbb{N}(x) \rightarrow \varphi(x))$) (so we can do an induction on the lengths of proofs). \mathcal{L} might or might not contain the predicate "true."

Ordinarily, a soundness proof doesn't count for much, since unsound theories routinely prove their own soundness. In an even fight between Field's system and

classical logic, a soundness proof would be dismissed as a distraction. But it isn't an even fight. Field is the upstart, trying to make a place for himself in a field in which classical logic has enjoyed nearly complete hegemony. His job is to convince people who have grown up with classical logic, are currently using it, and are inclined to regard it as tried and true, that they could do better. The soundness proof disrupts his sales campaign by giving potential customers a reason—by standards of reason they still regard as reliable—to think that the new product isn't an improvement, but, in fact, in many ways a downgrade. Take an arbitrary sentence θ . We may have no inkling whether θ is provable in the old logic or in the new one, and we may feel no inclination either to believe θ or to deny it. Nevertheless, we can prove the following, using methods that have served us well up till now: If the two systems disagree about θ , then the old system proves that θ , the new system doesn't prove that θ , and θ . Wherever the two systems disagree materially, the old system has the upper hand.

Field responds by attacking the theorem. This is quite an audacious move—the instances of the soundness schema are, after all, *theorems*; see Mostowski (1952)—but Field is a bold and resourceful thinker. His argument proceeds in two parts. First, a consistent, finitely axiomatizable system that includes basic arithmetic can't prove all the instances of the soundness schema. Where γ is the conjunction of the axioms, one of the instances of the schema is $(\ulcorner \neg\gamma \urcorner$ is a theorem of pure logic $\rightarrow \neg\gamma)$. If γ could prove this, it could prove its own consistency, contrary to the Second Incompleteness Theorem. Second, the restriction to finitely axiomatizable theorems is no serious limitation. For this, he appeals to a theorem of Craig and Vaught (1959), who, refining an earlier result of Kleene (1952a), show that any recursively axiomatized theory that lacks finite models, in a language built from a finite vocabulary, has a finitely axiomatized conservative extension. Field uses their result to show that, if you carefully put together just the wrong combination of language and theory, you'll be unable to prove soundness.

I am not persuaded. Fix the language \mathcal{L} . As long as you have, within \mathcal{L} , the capability of carrying out basic mathematical arguments, you'll be able to produce the soundness proofs. "Basic mathematical arguments" include proofs by mathematical induction. The soundness argument is an induction on the lengths of derivations. The finitely axiomatized theory constructed by Craig and Vaught provides only a restricted version of the principle of mathematical induction, so it lacks the resources to carry out the soundness proofs. What this shows is that the classical mathematician will be unable to prove the instances of the soundness schema if she is denied the use of her most powerful tool. That tool is mathematical induction.

Within Kripke's theory, in all its versions, there is a deep divide between what we want to say and what we can say, assuming we restrict ourselves to asserting what is true. We want to say that the liar sentence is neither truth nor false. That idea is a fundamental motivation for the system, and indeed we can prove within the classical metatheory that, within a fixed point, λ will be in neither the extension nor the antiextension of "T." We can formulate this judgment formally: $\neg(T(\ulcorner \lambda \urcorner) \vee T(\ulcorner \neg\lambda \urcorner))$. But even though we can formalize and prove it, working from the outside within a classical metalanguage, the judgment is not true, if the mark of

truth is membership in the extension of “T” in a preferred fixed point. Thus there is disturbing sense in which Kripke’s metatheory is self-refuting, proving judgments it tells us are untrue.

Field also relies on a fixed-point construction, more intricate than Kripke’s, so one could make the same complaint against him. But this time the attack would misfire. The difference is Field’s logical conception of truth.

The way a linguist understands her job, the truth conditions of the sentences of a language are established by speaker usage, and the semanticist’s task is to learn from the speakers what those truth conditions are. The truth about truth conditions is out there, waiting to be discovered. The special case in which the language being investigated is the linguist’s own idiolect presents peculiar epistemic opportunities and methodological challenges. The gravest of these methodological difficulties is the liar paradox. Among the features of usage the linguist is trying to describe is her own use of the word “true” as part of her technical jargon. Thus her theorizing modifies the very usage it is trying to describe. In such circumstances, a mismatch between the usage the theorizing creates and the usage the theory describes may well be unavoidable. In any event, the disparity between what one can see about the notion of truth the Kripke construction provides, looking at it from the outside, and what one can say about it from the inside, restricting one’s assertions to what the theory treats as true, shows that Kripke hasn’t succeeded in avoiding it.

The logical conception of truth is unashamedly egocentric, and it doesn’t aim to describe anything. There is no notion of a semantic theory trying and failing to describe what’s really true, independent of our theorizing. Instead, the aim of truth theory is to provide a way of using the word “true,” within one’s own idiolect, that is useful and coherent. Once one has a serviceable way of using “true” within one’s own language, one can extend it to other languages by translation, but that extension is not a part of the logical conception of truth, but an application of it.

From a logical point of view, the only formal constraint on a proposed way of using “true” is conservativeness, and the point of the fixed-point construction is to prove conservativeness. There is no presumption that the sentences that are assigned the value one by the fixed-point construction are the sentences that are really true.

The fixed-point construction shows that classical ω -logic has a conservative extension that contains a transparent truth predicate, that includes a binary operator that behaves in many of the ways we’d expect a conditional to behave, and that respects the strong Kleene rules for the traditional connectives. The good behavior that is being required of the conditional isn’t codified in explicit rules. Instead, the system of rules is specified as whatever system is shown to be conservative by the fixed-point construction. This highly indirect method of specifying the deductive system is unfortunate, in a way, since it means that the only way to see whether a logical principle counts as valid is to attempt, laboriously and without any assurance of success, to determine whether there is a model of the fixed-point construction in which it fails. A more explicit exhibition of the permitted methods of inference would have been welcome, although a complete deductive system is out of reach; the set of logically valid sentences is, as we shall see in an appendix, complete Π_2^1 .

An argument is said to be \aleph_κ -valid iff, whenever the fixed-point construction is erected over a classical ω -model of cardinality \aleph_κ or smaller, the argument

preserves the property of having value one in the fixed point. An argument is *valid* iff it's \aleph_κ -valid, for every κ . One can use the Löwenheim–Skolem theorem to show that an argument with finitely or denumerably infinitely many premisses is valid iff it's \aleph_0 -valid; see the appendix.

A common embarrassment for purveyors of nonclassical logics is their employment of classical logic in the metalanguage. If classical logic is so bad, we want to ask, how come you're using it in your own metatheory? This question is not an embarrassment for Field, who has a ready answer: The metatheoretic construction is part of pure mathematics, and within the language of pure set theory we can reason classically, for there we have the law of the excluded middle, $(\varphi \vee \neg\varphi)$.

The two questions—"Should we regard truth as a logical notion or a linguistic notion?" and "Should our logic be classical or not?"—are independent, and it's possible for someone with a logical conception of truth to hold onto classical logic. From this point of view, the purpose of strong supervaluational version of Kripke's fixed-point construction is to vindicate a certain system of rules by proving conservativeness. These rules include all the rules of the classical predicate calculus, and they include four special rules for truth⁷: T-Introduction (From φ , you may infer $T(\ulcorner\varphi\urcorner)$), T-Elimination, \neg T-Introduction, and \neg T-Elimination, but these come with an important caveat. Let's say a *restricted conditional proof* is one in which the word "true" does not appear within the assumed hypothesis. Then the four truth rules can be applied within direct proofs and restricted conditional proofs, but they cannot be employed within arbitrary conditional proofs. In addition, we have the axioms: "Every logical consequence of true sentences is true," "Every sentence is either true or false," and "No sentence is both true and false." We have the T-sentence $(T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi)$ whenever "T" doesn't appear within φ , but we don't have $(T(\ulcorner\lambda\urcorner) \leftrightarrow \lambda)$. We do, however, have $(\lambda \leftrightarrow \lambda)$, which means we don't have transparency.

We have a choice. We can allow ourselves full classical logic and restrict transparency, so that we can substitute $T(\ulcorner\varphi\urcorner)$ and φ when φ doesn't contain "T," but we cannot do so generally.⁸ Or we can uphold full transparency and restrict the logical rules that don't involve "T," so that we are only allowed the full range of classical inferences when the sentences involved don't contain "T." As far as I can see, we can choose whichever option best serves our purposes, and it might happen that different people make different choices because they have different purposes.

The classical option has the merits of simplicity and familiarity. These two advantages often go together, because a familiar system will seem simple because we've learned it well enough that we are able to use it with little effort. But even when we make allowance for the fact that new systems almost always seem complicated, Field's logic of conditionals is inordinately complex. (His logic of truth is as simple as can be.) The difficulty in using the system is greatly mitigated

⁷ See Friedman and Sheard (1987).

⁸ I wanted to say that the first option is one in which we *keep* classical logic, but this is disputable. One might say instead that the option consists in *extending* classical logic to encompass sentences that contain "T."

by the fact that, whenever we are dealing with sentences for which we can assume the law of the excluded middle, which include all the sentences that don't contain "T," we can employ plain classical logic.

Both options have an easy time with the very simple uses of the notion of truth as a device for blanket endorsement. They both get, "If everything the Pope says is true, then if the Pope says 'Penguins dream of flying,' penguins dream of flying." The classical logic approach can't get, "If everything the Pope says is true, then if the Pope says 'Everything the Grand Mufti says is true,' everything the Grand Mufti says is true," although it can infer, "If the Pope says 'Everything the Grand Mufti says is true,' everything the Grand Mufti says is true" from "Everything the Pope says is true." Field's approach can't get, "If the Pope is speaking truly, then if the Grand Mufti is speaking truly, they both are speaking truly," although it can infer, "If the Grand Mufti is speaking truly, then he and the Pope are both speaking truly" from "The Pope is speaking truly." Both systems get stuck trying to advance from the validity of an inference to the validity of the corresponding conditional, the classical approach because the application of the rules for truth within conditional proofs is restricted, and Field's approach because it doesn't allow conditional proof except where it has excluded middle.

We are offered a bargain. We can obtain full transparency, but the price is that, when we are reasoning with the notion of truth, we have to give up the easy and comfortable familiarity of classical logic. To my way of thinking, the deal isn't worth the price, but in saying this I am reminded of a remark of Paul Feyerabend (1975, Chap. 12). It was a foregone conclusion that, in their early confrontations, the Aristotelian theory would get the better of Copernicanism. The Aristotelians had had centuries to integrate their astronomical theory into a systematic, comprehensive worldview, whereas the rival account was still in its early formative stages. It was, Feyerabend said, like putting up an infant to fight a grown man. The situation is similar here. We are comparing a venerable and well-established approach to logic to a brand new system by asking which system has the more impressive list of useful applications. Of course the newcomer is going to have the shorter list. As the system matures, it will rack up more accomplishments. I wouldn't be surprised if it grows to become a formidable adult.

Appendix on validity

We want to show that, for arguments with countably many premises, \aleph_0 -validity coincides with validity. Then we'll use the proof to determine the complexity of the set of valid sentences.

To say that the argument from a countable premiss set Γ to φ is invalid is to say that there is a fixed-point model with standard integers in which all the members of Γ have value 1 but φ does not. This can be formalized as a Σ_1 sentence of the language of set theory. If it's true, then it's true in some model of the form $\langle V_\lambda, \in \rangle$, where λ is a strong limit cardinal. Taking the transitive collapse of a countable elementary submodel, we'll still get a fixed-point model with standard integers in which all the members of Γ have value 1 but φ does not, only now the

fixed-point model will be countable. The existence of such a model shows that the inference is not \aleph_0 -valid.

The same argument shows that, for any α , an argument with \aleph_α or fewer premisses is valid iff it's \aleph_α -valid.

Turning to examine the set of valid sentences, it is clear that a complete axiomatization is out of the question. For a theory of truth to be any use at all, it needs to be able to identify sentences syntactically. If we are encoding the syntax by Gödel numbers, having a syntactic theory is tantamount to having an arithmetical theory. To avoid fretting over irrelevant pathologies that arise in models that misrepresent the syntax, Field lays it down that the only models to be considered are ones that have standard integers. This ensures the set of set of valid sentences will be at least as complex as the set of arithmetical truths. It turns out that it's significantly more complex, specifically, complete Π_2^1 .

φ is invalid iff there is a well-founded model of Zermelo set theory⁹ in which there is a fixed-point model that doesn't assign φ the value 1. The Löwenheim–Skolem argument shows that this happens iff there is a well-founded model of Zermelo set theory with domain the set of natural numbers in which there is a fixed-point model in which φ is assigned a value different from 1. This can be formalized as a Σ_2^1 arithmetical statement.¹⁰

We need to show that every Π_2^1 set of natural numbers is 1-reducible to the set of valid sentences. For i a natural number and A a set of natural numbers, define the *ith tree in A* to be the set of all finite sequences x with property that x and all its initial segments are in the set whose characteristic function is calculated by the *ith* oracle Turing machine with an oracle for A , or to be the empty set if the machine doesn't calculate a characteristic function. The set of numbers i such that there is an infinite path through the *ith* tree in A is a complete Σ_1^1 -in- A set,¹¹ so $\{i: \text{for each set } A, \text{ there is an infinite path through the } i\text{th tree in } A\}$ is a complete Π_2^1 set. We want to show that this set is 1-reducible to the set of valid sentences.

In the language obtained from the language of arithmetic by adding, in addition to the truth predicate “T” and the new conditional “ \rightarrow ,” a new monadic predicate “R,” form, using Gödel's self-referential technique, a formula $\sigma(i, \underline{x})$ equivalent to:

$$x \text{ is in the } i\text{th tree in the set of natural numbers that satisfy } R \wedge (T(\ulcorner \sigma(i, x) \urcorner) \rightarrow (\exists y) (y \text{ is a member of the } i\text{th tree in the set of natural numbers that satisfy } R \wedge y \text{ extends } x \wedge T(\ulcorner \sigma(i, y) \urcorner))).$$

Where $\langle \rangle$ is the empty sequence, we want to see that $\sigma(i, \langle \rangle)$ is valid iff, for each set A , there in an infinite path through the *ith* tree in A .

Suppose that $(\forall A)(\text{there is an infinite path through the } i\text{th tree for } A)$. Take a fixed-point model, and let A be the set of natural numbers the model places into the extension of “R.” We know that there is an infinite path through the *ith* tree in A . We

⁹ Zermelo set theory is like standard set theory except that it doesn't have the replacement principle. For our purposes, employing such a very strong theory is overkill, but it gets the job done.

¹⁰ We focus our attention on sentence validity, that is, on the validity of inferences without premisses, for simplicity. The same argument shows that every Π_2^1 set of premisses has a Π_2^1 set of consequences.

¹¹ See Rogers (1967, §16.3). Our terminology follows his.

want to see that, for that every node x along the path, $\sigma(\underline{i}, x)$ is assigned the value 1 by the model; this will show that $\sigma(\underline{i}, < >)$ is assigned the value 1 by the model. We know that $\sigma(\underline{i}, x)$ can't be assigned the value 0, because that would give us a conditional with value 0 whose antecedent had value 0, which is impossible. The possibility we have to worry about is that $\sigma(\underline{i}, x)$ has value $\frac{1}{2}$. If $\sigma(\underline{i}, x)$ has value $\frac{1}{2}$, then we have a $\frac{1}{2}$ -valued conditional with a $\frac{1}{2}$ -valued antecedent, so the consequent must have value 0 or $\frac{1}{2}$. If we take z to be a node on the path further along than x , $\sigma(\underline{i}, z)$ can't have the value 1, since if it did " $(\exists y)$ (y is a member of the i th tree in the set of natural numbers that satisfy $R \wedge y$ extends $x \wedge T(\ulcorner \sigma(\underline{i}, y) \urcorner)$)" would have value 1. It can't have value 0, for the same reason $\sigma(\underline{i}, x)$ can't have value 0. So $\sigma(\underline{i}, z)$ must have value $\frac{1}{2}$, which implies that " $(\exists y)$ (y is a member of the i th tree in the set of natural numbers that satisfy $R \wedge y$ extends $x \wedge T(\ulcorner \sigma(\underline{i}, y) \urcorner)$)" has value $\frac{1}{2}$. Where Δ is an acceptable ordinal, both the antecedent and the consequent of the conditional " $(T(\ulcorner \sigma(\underline{i}, x) \urcorner) \rightarrow (\exists y)$ (y is a member of the i th tree in the set of natural numbers that satisfy $R \wedge y$ extends $x \wedge T(\ulcorner \sigma(\underline{i}, y) \urcorner)$))" have the value $\frac{1}{2}$ at level Δ , which means that the conditional, and hence its antecedent and consequent, will assume the value 1 at level $\Delta+1$. Once these sentences achieve the value 1, they'll maintain the value 1 at every subsequent level, with z serving as a witness to the consequent's existential claim. So the ultimate value of $\sigma(\underline{i}, x)$ is 1, contrary to hypothesis.

Now for the other direction, suppose that there is a set A for which there is no infinite path through the i th tree for A . This means that the partial order on the nodes of the tree that we get by stipulating that y is less than x iff y extends x is well-founded, so that we can do inductions. We want to show that, for every node x of the tree, $\sigma(\underline{i}, x)$ is assigned a value different from 1 in the fixed-point model obtained by taking the ground model to be the natural number system and taking A as the extension of " R ." We may assume the tree is nonempty, since otherwise $\sigma(\underline{i}, x)$ has value 0, for every x . Suppose that x is a node of the tree and that, for every y below x , $\sigma(\underline{i}, y)$ is assigned a value different from 1. Then the consequent of the conditional " $(T(\ulcorner \sigma(\underline{i}, x) \urcorner) \rightarrow (\exists y)$ (y is a member of the i th tree in the set of natural numbers that satisfy $R \wedge y$ extends $x \wedge T(\ulcorner \sigma(\underline{i}, y) \urcorner)$))" has a value different from 1. If $\sigma(\underline{i}, x)$ had value 1, we'd be assigning the value 1 to a conditional whose antecedent had value 1 and whose consequent had a value different from 1, which is absurd. It follows by induction that the value of $\sigma(\underline{i}, < >)$ is different from 1.¹²

Bibliography

- Barwise, K. J. (1975). *Admissible sets and structures*. Berlin: Springer-Verlag.
- Beall, J. C. (2009). *Spandrels of truth*. Oxford: Oxford University Press.
- Craig, W., & Vaught, R. L. (1959). Finite axiomatizability using additional predicates. *Journal of Symbolic Logic*, 23, 298–308.

¹² By a sophisticated analysis, Philip Welch (2008) has determined the recursive-isomorphism type of the set of ultimate truths obtained by erecting the fixed-point construction over the standard model of arithmetic. These ultimate truths will be the consequences of " $(\forall x)\mathbb{N}(x)$ " that contain no non-arithmetical predicates other than " T ."

- Feferman, S. (1984). Toward useful type-free theories, I. *Journal of Symbolic Logic*, 49, 75–111. Reprinted in Martin (1984), pp. 237–287. Page references are to the reprint.
- Feyerabend, P. (1975). *Against method*. London: Verso.
- Field, H. (1972). Tarski's theory of truth. *Journal of Philosophy*, 69, 347–375. Reprinted with a postscript in Field (2001), pp. 3–29. Page references are to the reprint.
- Field, H. (2001). *Truth and the absence of fact*. Oxford: Oxford University Press.
- Field, H. (2003). A revenge-immune solution to the liar paradox. *Journal of Philosophical Logic*, 32, 139–177.
- Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.
- Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 40, 1–10.
- Kleene, S. C. (1952a). Finite axiomatizability of theories in the predicate calculus using additional predicates. *Memoirs of the American Mathematical Society*, 10, 27–68.
- Kleene, S. C. (1952b). *Introduction to metamathematics*. New York: Van Nostrand.
- Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72, 690–716. Reprinted in Martin (1984), pp. 53–81.
- Leeds, S. (1978). Theories of reference and truth. *Erkenntnis*, 13, 111–129.
- Martin, R. L. (Ed.). (1984). *Recent essays on truth and the liar paradox*. Oxford: Oxford University Press.
- Moschovakis, Y. N. (1974). *Elementary induction on abstract structures*. Amsterdam: North Holland.
- Mostowski, A. (1952). On models of axiomatic systems. *Fundamenta Mathematicae*, 33, 133–158.
- Putnam, H. (1965). A philosopher looks at quantum mechanics. In R. C. Colodny (Ed.), *Beyond the edge of certainty* (pp. 75–101). Englewood Cliffs, NJ: Prentice-Hall. Reprinted in Putnam (1975, pp. 130–158).
- Putnam, H. (1975). *Mathematics, matter, and method. Philosophical papers* (Vol. 1). Cambridge: Cambridge University Press.
- Quine, W. V. O. (1984). *Philosophy of logic* (2nd ed.). Cambridge, MA: Harvard University Press.
- Rogers, H. (1967). *Theory of recursive functions and effective computability*. New York: McGraw-Hill.
- Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1, 261–405. English translation by J. H. Woodger in Tarski (1983, pp. 152–278).
- Tarski, A. (1936). O ugruntowaniu naukowej semantyki. *Przeegląd Filozoficzny*, 39, 50–57. English translation by J. H. Woodger in Tarski (1983, pp. 401–408). Page references are to the translation.
- Tarski, A. (1983). *Logic, semantics, metamathematics* (2nd ed.). Indianapolis: Hackett.
- van Fraassen, B. C. (1966). Singular terms, truth value gaps, and free logic. *Journal of Philosophy*, 63, 464–495.
- Welch, P. (2008). Ultimate truth vis à vis stable truth. *Review of Symbolic Logic*, 1, 126–142.