Philosophical Studies (2005) 124: 71–104
DOI 10.1007/s11098-004-4912-z

V. McGEE

TWO CONCEPTIONS OF TRUTH? – COMMENT

1. TRUTH AND MASS

Hartry Field (1986 and 1994) has suggested that certain conflicts within our ordinary usage of the word "true" are best resolved by allowing that there are two conceptions of truth, each legitimately useful in its own place: the "disquotational" or "deflationary" conception, which regards the (T)-sentences[1] as true by definition, merely in virtue of the meaning of the word "true"; and the "correspondence" conception, which treats the truth conditions of sentences as effects of the linguistic practices of speakers.[2]

It is not Field's contention that the word "true," as it is ordinarily used, is simply ambiguous, the way "seal" and "bank" are. It is perfectly clear to ordinary speakers that they use the word "seal"[3] for two quite different sorts of things, and it's not that way with "true." The current status of the word "true" is similar to that of the word "mass" before relativity theory.[4] There were, as we recognize today, two different prerelativistic uses of the word "mass," although it seemed like a univocal usage to speakers at the time. Their usage was based on the presumption (although it's not the sort of assumption one would be aware of) that the mass of a thing doesn't depend of how fast it is moving. Rejecting that presumption, we now see "rest mass" and "inertial mass" where they saw only "mass."

To get to the conclusion that two notions of mass were in play, Einstein had to do more than to demonstrate that experimental procedures that were commonly thought of as "mass measurements" would, under suitable circumstances, yield different answers. That evidence would be compatible with a single quantity often mismeasured. Even if the measured values were sharply bimodal, that would be consistent with a

single quantity systematically mismeasured. One might say, for example, that "mass" always equals "rest mass," and that earlier procedures sometimes gave different answers because they neglected the fact that rapidly moving bodies appear heavier than they really are. To arrive instead at the conclusion that there are two distinct, physically significant quantities, Einstein had to embed the two notion in a theory, where each, in its separate place, had something worthwhile to contribute.

One can always employ conceptual bifurcation as a response to apparent conflict, but doing so is not, as a general rule, a wholesome approach. Whereas conceptual mitosis is sometimes appropriate, as the "mass" example clearly shows, it is a response we should make with reluctance, lest we dull the edge of dialectic. At the end of the first book of the *Republic*, Socrates might have assuaged Thrasymachus' temper by allowing that there are several different conceptions of justice and that Thrasymachus has elucidated one of them. Socrates does not do so. Instead, he rejects Thrasymachus' proposal outright, thus sparing us the Ashcroftic conceit that there are two standards of justice, justice-as-the-advantage-of-the-stronger, appropriate to times of war, and a less muscular standard for peaceful times.

To justify the two concepts of truth, one needs to identify a distinctive load that each of the new concepts can usefully bear. Field's paper has been much discussed and (rare for a philosophy paper) widely accepted, and more-or-less standard strategies for providing justifications have emerged. For the correspondence conception, which sees the truth conditions of sentences as effects of the activities of a community of speakers and causes of their verbal behavior, the strategy is obvious, at least in broad outline. We can't understand the things people do with their mouths, their pencils, and their keyboards, or the ways these activities affect the other things people think and do, unless we suppose that the sounds and marks the people make constitute meaningful speech, and the most likely way to understand meaning is in terms of truth conditions. I shall endorse this strategy, although not without some grumbling about how faintly the purported explanations are sketched.

The justification for employing the disquotational notion, which understands the truth of the sentence "Snow is white" as purely an effect of the whiteness of snow, not dependent on our usage of the word "snow," has to be more subtle, since it's part of the disquotationalist story that the notion of truth is unfit to play a significant theoretical role in causal explanations. The standard story has it that truth functions as a logical device, so that we can use the sentence, "Everything the Pope says *ex cathedra* is true," to express an infinite conjunction of sentences of the form, "If the Pope says *ex cathedra* that φ, then φ." I am not entirely satisfied with this explanation. This way of employing the notion of truth, while genuine, is not as valuable as it's cracked up to be, because we can accomplish practically the same things using correspondence truth. Instead, I shall recommend paying attention to the cognitive, rather than the communicative, role of disquotational truth, with particular emphasis on its role in mathematics. Introducing the word "true," disquotationally defined, and permitting it to appear within axiom schemata enables us to discover new mathematical truths.

A different usage of disquotational truth arises from the thinness of the broth the correspondence theorist is offering. In their theoretical role, the two conceptions of truth differ sharply, but (for our own language) they differ in extension scarcely at all. We know very little about correspondence truth. We know that our usage has to fix the truth conditions of our sentences somehow, but we have only the sketchiest ideas how it is done. On the other hand, we know a lot about which sentences are disquotationally true, so the fact that the two notions are nearly coextensive is an important datum toward the construction of a correspondence theory. Knowing which sentences are disquotationally true won't give us the explanations the correspondence theorist seeks, but it will help us identify the phenomena to be explained. One reason the disquotational notion of truth is valuable is that it gives us a extensionally good approximation to correspondence truth. As I understand it, the two-conceptions thesis requires more than

this. It requires that disquotational truth be valuable in its own right.

2. THE DISTINCTION

The paradigm, at least in my mind, of a correspondence theory of truth is the program Field sketches, at the end of (1972), of taking the account Tarski gives in (1935) and supplementing it with a causal theory of reference. It should be said at once that the program doesn't work. It suffers from two problems, one readily correctable, the other not. The first objection, urged by Scott Soames (1984) and Robert Stalnaker ((1984), pp. 30f), is that the problems Field astutely diagnoses with Tarski's treatment of the nonlogical primitives are also to be found in Tarski's treatment of the logical connectives, about which Field makes no complaint. Soames and Stalnaker are right, but the problem can be fixed by further supplementing Tarski's theory, explaining how the semantic values of the logical connectives are provided by their inferential roles.[5] The harsher problem, as Field himself soon realized (see Field (1974)), is that the program founders on the shoals of the inscrutability of reference. Perhaps some variant of Field's program can be salvaged, but the plan in its original form cannot.

We now see that the program of "Tarski's Theory of Truth" is a product of wishful thinking, and precisely for that reason it is useful, for it reveals to us what we'd really like a correspondence theory to look like, if we could get whatever we wanted. What we'd really like is a straightforward causal, historical explanation of how the truth conditions of sentences are fixed by the practices of speakers, with no metaphysical funny business.

The name "correspondence theory" seems a little inapt. The program outlined in "Tarski's Theory of Truth" is a correspondence theory *par excellence*, yet we don't find anything in it about a highly elaborated *ad hoc* edifice of facts, nor do we find a mysterious relation of correspondence between sentences and facts. The reference relation for simple terms is a bit mysterious if we leave things where Tarski left them – that's the main point

of Field's paper – but we don't have to leave things where Tarski left them. Apart from a relatively harmless foundation of syntax and core mathematics, truth attributions got by Tarski's theory don't have any ontological commitments beyond those of the object theory. Field's supplement has further commitments, requiring us to acknowledge social processes by which a community's linguistic practices forge a causal connection between words and their referents. There are considerable doubts, dating from Chapter 2 of *Word and Object*, whether Field's supplement can get the job done, but Quine's misgivings aren't ontological. There is nothing metaphysically extravagant in the program Field outlines.

The name "correspondence theory" is objectionable because the paradigm correspondence theory doesn't say anything about facts or correspondence. The epithet "deflationary conception" as a name for its rival is exceptionable for the same reason. The name "deflationary conception" suggests that its opposite is inflationary, and there is nothing metaphysically inflationary about the program Field proposes. Quite the contrary, "Tarski's Theory of Truth" is honest, plainspoken physicalism. The alternative designation, "disquotational conception," which is due to Quine (1970, p. 12), is right on target. On the disquotational conception, adding "is true" to the quotation name of "Snow is white" cancels the effect of the quotation marks.

Having complained about the names given to the distinction, it is now time to grouse about the way the distinction is drawn. In (1972), it is drawn modally. According to the disquotational conception, the (T)-sentences are necessary. This has the consequence that, if we had used the word "white" so that it applied to things that are warm, "Snow is white" would still have been true. On a correspondence conception, the (T)-sentences are true only contingently.

I don't think the modal characterization quite succeeds in getting at the distinction we want. On a correspondence conception, there are two different readings of the (T)-sentence for "Snow is white," on one of which the (T)-sentence is necessary. To see this, we rely on two ideas, both loosely attributable to

David Kaplan (1989). First, the correspondence truth or falsity of an utterance is a product of two factors, facts about linguistic usage and the circumstances of the utterance that determine the truth conditions of the utterance, and facts about what the world is like that determine whether those conditions are met, and it is possible to vary these factors independently. Second, it is possible to freeze one of the factors by employing an "actually" operator. In asking, "If we had used the word 'white' so that it applied to things that are warm, would 'Snow is white' have been true?" it is natural to understand "true" to mean "true in our language as we would have been using it," that is, true in the language as we use it in the world most like the actual world in which we apply the word "white" to things that are warm. So understood, the answer to the question is clearly "No," not only for the correspondence theorist, but for Field's so-called "moderate" disquotationalist, who is willing to extend the notion of truth to languages different from her own by translation. However, if we insert the word "actually," asking "If we had used the word 'white' so that it applied to things that are warm, would 'Snow is white' have been true in our language as we actually use it?" the answer will be "Yes," for the correspondence theorist as well as for the disquotationalist. Necessarily, "Snow is white" is true in our language as we actually now use it if and only if snow is white. With proper attention to the fact that the truth of a sentence is relative to a language and a context,[6] both the necessary and the contingent reading of the (T)-sentences are available to both the disquotationalist and the correspondence theorist.

The real battleground is what Field (1994a) calls "factually defective" utterances: borderline applications of vague terms (on almost all accounts); sentences containing nondenoting names (on many accounts); conditionals (on a number of accounts, notably Adams (1975)); moral and aesthetic judgments (according to emotivists); theoretical judgments (for radical empiricists); and so on. Here I shall focus attention almost entirely on the application of vague terms, like "poor." According to correspondence theory, in order for an utterance

of the sentence "Clare is poor" to be true, the way members of our community use the word "poor," together with features of the context of utterance that determine what standard of poverty is relevant to the situation, together with Clare's financial situation, have to make it true. Likewise, in order for an utterance of the negation, "Clare is not poor," to be true, the same features have to make the negation true, in which case we shall say that the situation is one in which "Clare is poor" is false. It is implausible that our apparently easygoing use of the word "poor" is actually so rigorous that it establishes, for each possible situation, an exact, down to the last penny, partition of people into those who satisfy "poor" in the situation and those who satisfy "not poor" in the situation. The credible hypothesis is that our usage leaves some cases unsettled, so that there are circumstances in which it wouldn't be truthful, as the correspondence theorist reckons truth, to say either "Clare is poor" or "Clare is not poor," though either statement would be meaningful. For the correspondence theorist, there are meaningful assertions that are neither true nor false.

For the disquotationalist, repudiating bivalence is not an option, for the principle is built into the very meaning of the word "true." The (T)-sentence, "'Clare is poor' is true if and only if Clare is poor" is, on the disquotational account, not merely true, but, if I may use the word, analytic. Putting it this way presumes that the same standard of poverty is being applied in the quoted and unquoted occurrences of "Clare is poor." Whenever we say the following meaningfully, we say something true, according to the disquotationalist:

"Clare is poor" is true in the present context if and only if Clare is poor.

Likewise,

"Clare is not poor" is true in the present contest if and only if Clare is not poor.

Consequently, by classical logic, within the present context (assuming it is one in which "Clare" has a definite referent), either "Clare is poor" or "Clare is not poor" will be true, no matter what Clare's financial situation.

Here is the most dramatic difference between the disquotational and correspondence conceptions of truth: The former is committed to bivalence, even for factually defective utterances. The latter is not.

3. THE NEAR-INVISIBILITY OF THE DISTINCTION

It took an Einstein to recognize the two conceptions of mass, because the distinction is so subtle. Experimentally, the discrepancy is so tiny that no one would have noticed it without a theory directing where to look, a theory that required profound insights into the nature of space, time, and matter. It requires no similarly profound insight to recognize the appearance of meaningful but factually defective discourse. In saying this, it is certainly not my intention to deny that deep thought was required to construct an emotivist ethics or to devise Adams' probabilistic theory of conditionals. But some factual defects, notably the indeterminacy that arises at the border of a vague term, are apparent to anyone. Vagueness is ubiquitous throughout natural language, and it requires no special talent or training to recognize it. Moreover, the correspondence theorist's response to vagueness, which is to deny that the borderline attributions are either true or false, is something that would naturally occur to anyone. That being so, why was Field's acknowledgment that there are two conceptions of truth a deep insight, rather than a philosophical commonplace? Why isn't the distinction something we've known about for ages?

Part of the explanation is the simple fact that nobody will ever be able to identify an utterance that is true on one conception but not on the other. The most we can hope to do is to identify pairs of sentences with the property that, with respect to a particular context, one or the other is disquotationally true but neither is correspondence-true. If Clare is in appropriately straitened circumstances, "Clare is poor" and "Clare is not poor" form such a pair. We know by logic that either Clare is poor or Clare is not poor, which entails, according to a disquotational conception of the meaning of the word "true," that

either "Clare is poor" or "Clare is not poor"is true. It may nonetheless happen that the totality of facts about usage and facts about Clare's financial circumstances fail to determine an answer to the question, "Is Clare poor?"

If Clare is a borderline case, even an omniscient being will be unable to answer the question, "Is Clare poor?" God knows exactly how much money Clare has, of course, and He knows precisely how Clare's money and possessions compare with those of her neighbors. He also know how every English speaker uses the word "poor." But it could happen that all this knowledge doesn't add up to a definite answer. You might be inclined to say that, if Clare is poor, then God knows that she is poor, and if Clare isn't poor, God knows that she isn't poor, so that, in either case, God knows whether Clare is poor, but saying this overlooks the fact that divine knowledge is being expressed in human language. The practices of mortal speakers determine the conditions of application of the word "poor,"and the perfection of divine knowledge doesn't obviate the imprecision of human language. Even God doesn't know whether the smallest large number is divisible by seven.[7]

With respect to simple judgments arrived at more-or-less directly, the maxim, "Assert things that are true and avoid asserting things that aren't," will sanction the same speech acts whether "true" is understood disquotationally or in terms of correspondence. Does the same hold for complex statements or simple statements reached by complex inferences? If we take it as common ground between the two camps that a disjunction is true if and only if one or both disjuncts are true, then the answer is definitely "No," since, if Clare is on the border, the correspondence theorist will be forced to declare, "Clare is poor or Clare is not poor," neither true nor false. More generally, the correspondence theorist will say that speakers reason badly if they employ classical logic within a discourse in which there are actual or suspected factually defective sentences. As Russell puts it (1923, p. 65), "All traditional logic habitually assumes that precise symbols are being employed. It is therefore not applicable to this terrestrial life, but only to an imagined celestial existence."

This is an unwelcome conclusion. Vagueness appears everywhere in natural language. One can make a case that mathematical symbols are precise enough to equip each sentences of pure mathematics with a uniquely determined truth value, but outside pure mathematics, vagueness is ubiquitous. It's hard to think of even one nonmathematical predicate that doesn't have actual or hypothetical borderline cases. Moreover, we constantly and unashamedly employ classical logic and classical mathematics in reasoning with concepts with imprecise boundaries. Because of the vagaries of immigration status, the set of people who live in Canada isn't a sharply defined set and the population of Canada isn't a sharply defined number, and in spite of that demographers studying Canadian social trends make unabashed use of classical statistics. They don't worry that maybe statistical theory doesn't apply when you're doing sociology.

The size of the stake here can seem less than it really is, if we enforce an artificial boundary between logic and mathematics. We seldom have occasion to make claims of the form "φ or not-φ," and if a philosophical analysis forces us to foreswear such assertions, it doesn't seem like such a sacrifice. But logic and mathematics form a whole, and you can't cripple the former and leave the latter unharmed. If we insist that an existential sentence is untrue if each of its instances are untrue (on a reading that treats borderline attributions as neither true nor false), then it will be untrue that the population of Canada is less than fifty million, and a true demographic analysis will be out of reach. You can't do demography if you can't apply the concept of number to social classes as well as classes of numbers.

Classical applied mathematics is the most successful scientific endeavor of all time. Nothing else comes even close. To ask that we abandon the endeavor to satisfy the demands of philosophical semantics strikes me (and struck me even before I read David Lewis's "Credo"[8]) as nothing short of ludicrous.

Fortunately, there's no need. The practices of the community of speakers determine the meanings of the words of the language, which determine the truth conditions of the sentences,

but there's no reason to presume that this determination proceeds by a direct compositional semantics, which treats a disjunction as true only if one or the other disjunct is true. Field (1977) has argued persuasively for the indispensable part played by conceptual role in meaning fixation. In particular, the meaning of the logical connectives is established by the inferential role of those connectives.[9] If speakers reason classically, their disposition to reason that way is built into the meaning of the word "or." We can ask of an individual speaker whether she uses the connectives in logically legitimate ways, as a way of asking for assurance that her usage conforms to community standards. But inferential practices of the whole community vindicate themselves, in the sense that community practices establish the meanings of the connectives, and the meanings of the connectives justify the inferential practices.[10]

Correspondence truth still has compositional semantics, even when there are factual defects, but the compositionality is indirect. Speakers' practices single out, not a unique intended interpretation of the language, but a family of acceptable interpretations,[11] and truth in an interpretation is compositionally determined, by the method of Tarski (1935). The good inferences are the ones that preserve truth in an acceptable model, and by this standard all the classical inferences are legitimate.[12]

For complex sentences as well as simple, the maxim, "Assert things that are true and avoid asserting things that are untrue," will sanction the same assertions whichever notion of truth is being employed, and, moreover, both notions legitimate the same logical inferences. In fact, the distinction isn't one speakers ever need to be aware of. The distinction is needed for an adequate account of meaning fixation, but you can speak a language perfectly well without a theory of meaning fixation.

Imagine a sequence of ten thousand ceramic tiles, gradually shading from red to orange. It follows logically from the fact the tiles are arranged in a sequence with the first tile red and the last tile not red that there is a last red tile. Even so, speakers, no matter how sharp-eyed, aren't able to point to a tile and say, "That tile is red, unlike the tile right after it." There is a range

of tiles for which speakers don't know and can't say whether the tile is red, no matter how carefully they examine it. There are strategies we might try for adjudicating the hard cases. We might try examining the tiles spectrographically, with the thought in mind that our classification ought to respect natural physical borders, even borders that aren't apparent to casual inspection. We might also try asking people. We might try asking painters, on the theory that they have learned to attend to colors more carefully than the rest of us, or we might try polling the public at large, with the idea of getting a better sense of what the community's standards of redness are. The tiles that resist all such efforts fall into three (possibly empty) categories. The first consists of tiles that are correctly classified as "red," even though the facts about the tile and about linguistic usage that make this a correct classification are too subtle to discern, even on a detailed investigation. The second consists of tiles that are correctly classified as "not red," for similarly subtle reasons. The third consists of tiles for which there is no fact of the matter, because the color of the tiles and the totality of facts about our usage of "red" fail to determine an answer. Which category a tile falls into is an effect of complex socio-linguistic forces that determine the range of correct application of an adjective, but it isn't, as far as I can see, a cause of much of anything. It makes no difference to how speakers talk about the tiles.

Timothy Williamson (1994) has argued trenchantly that the third category is empty. His arguments are many and ingenious, but their basic thrust is that *epistemism*, which holds that vagueness is entirely a matter of epistemic rather than semantic limitations, is the only way to relieve the tensions within our ordinary conception of truth that led Field to postulate that there were two concepts under a single name. Epistemism holds that speakers' usage determines, relative to a context, completely sharp boundaries for each meaningful term, and that appearances to the contrary are caused by the fact that the boundaries aren't discernible by speakers. Because it's meaningful to say of a person that she's not poor, really, but sort of down at the heels, usage must determine precise, down to the

last toothpick, upper and lower financial bounds for the range of application of the phrase "not poor, really, but sort of down at the heels." I find this doctrine incredible, but I have to admit that I don't have a well-confirmed theory of meaning fixation with which to rebut it.

4. THE MAIASAUR PROBLEM

The distinction between the two notions of truth was hard to see, but once we are aware of it, we need to make up our minds whether to keep both notions as part of our conceptual repertoire, or to reject one or the other as illegitimate. To justify the former course, we need to see, at least in outline, how each of them plays a useful role that cannot be filled by the other. For correspondence truth, the plan in straightforward. The notions of truth values and truth conditions are utilized in causal explanations of verbal behavior. They are complex because the phenomena to be explained are complex, but they aren't different in kind from other theoretical concepts of the social sciences.

On the chalkboard in my office, I have written the following sentence, copied from the website of the Royal Ontario Museum:[13] "Maiasaurs were highly social animals that traveled in herds of as many as 10,000." The appearance on the blackboard of this particular string of chalk marks is an effect of events that occurred some 70,000,000 years ago. It is an extremely subtle effect, depending not only on the herding behavior of maiasaurs but on the intimate details of their family life. The marks at the far left, a token of the word "maiasaurs," which means "good mother lizards," appear there because Dr. Jack Horner, who named the species, discovered, to great surprise, that adult maiasaurs cared for their young. It is also a highly indirect effect. The chalk marks do not resemble a dinosaur in size, shape, or smell, and there is no reason to suppose that either the chalk or the slate was ever in the vicinity of maiasaurs. Of course, virtually every event has distant effects, but that the social life of dinosaurs should have such distant and subtle effects *that we are able to discern* is, I am sure

you will agree, an astonishing fact. It is not, however, an extraordinary fact. We are able to discern similarly subtle and indirect causal chains routinely, issuing in such effects as chalk marks, ink blotches, and auditory disturbances, and originating in such diverse sources as the Big Bang, the rings of Saturn, and the Second Punic War.

The astonishing fact that there is an easily recognized connection between the dinosaurs and my chalk marks demands an explanation, and it deserves a better explanation than we get at the level of molecular chemistry: sunlight reflected from the maiasaurs' fossil remains stimulated Dr. Horner's optic nerves, causing digital muscle twitches later on as he sat at the keyboard. This pallid biomechanical explanation neglects something fundamentally important: human beings are highly social animals who communicate by language.

To understand the connections between our chalk marks and auditory outbursts and the things or states of affairs that they are about is a central problem of the cognitive sciences. In spite of its centrality, it hasn't an accepted designation, so I'll give it a name. I'll call it the *Maiasaur Problem*. The format that a solution is going to take is clear: The chalk marks form a meaningful English sentence; the meaning of the sentence, given to it by the activities of Dr. Horner and other speakers of the language, connects the sentence with the herding activities of dinosaurs. Not quite so obvious, but still the most likely hypothesis in town, is that meaning is to be understood, at least in part, in terms of truth conditions. The activities of speakers give the sentence truth conditions, which might or might not be satisfied, depending on the behavior of the dinosaurs.

To understand how a community's activities give their sentences truth conditions is surely a difficult task. Among other things, it incorporates one of the central mysteries of philosophy, how to derive "ought"s from "is"s. The things speakers actually say in different circumstances somehow or other give rise to norms that determine what speakers ought correctly to say on various occasions. The problem we confront here is not the full-fledged problem how Nature can start with atoms falling in the void and wind up with a world with norms in it,

since the norms involved are conditional: what ought we say in order to advance our conversational purposes? Even so, the problem is surely hard.

It's a hard problem, but we have a plausible starting point. I know what the sentence "Maiasaurs traveled in herds of as many as 10,000" would mean if I said it, and, since the authors of the website and I are members the same speech community, it is reasonable to surmise that what they meant was pretty much the same as what I would have meant. Using the abbreviated description of the fossil evidence found on the website, it is easy to construct a credible story leading from the bone pile to the electronic utterance of a sentence with that meaning.

This is the beginning of the story, but only the beginning. Conspicuously absent from it is an account of why my words mean what they do. As an English speaker, I am privileged to know what common English sentences mean, but I have no privileged access to English etymology. If we want an explanation of the causal process that leads from the dinosaurs to the inscription on the blackboard, we won't be satisfied with a story that reads, "And then something magical happens that invests the words with meaning." Something natural happens that invests the words with meaning, and we want to know what it is.

Folk semantics proposes an answer. Causal connections between speakers and objects link certain particular words and certain particular objects in such a way that, for example, "maiasaur" refers to maiasaurs, and to noting else. Once reference is established, truth conditions are imposed compositionally so that, for example, "Maiasaurs were warm-blooded" is true if and only if maiasaurs were warm-blooded, and false if and only if maiasaurs weren't warm blooded.

The intuitive appeal of this folk-semantic story is nearly irresistible, but powerful philosophical arguments compel us to resist it nonetheless. Beginning with Quine (1960), there have been a number of demonstrations that there isn't anything in the practices of a community of speakers that attaches unique referents to the speakers' terms. Peter Unger's "Problem of the Many"[14] is an especially persuasive variety of this species.

That's at the level of words. Indeterminacy persists up to the level of sentences. Folk semantics maintains that truth conditions are established by community usage in such a way as always to verify the (T)-sentences, but such a commonplace example as "Clare is poor" shows that this assumption cannot be maintained.

We need to replace folk semantics with an alternative explanation of the connection between the sentence "Maiasaurs traveled in herds of as many as 10,000" and the herds of maiasaurs. The currently most promising plan is an updated version of Field (1972): supplement Tarski's (1936) characterization of truth under an interpretation (as opposed to his (1935) characterization of truth) with an account of how the practices of the speech community establish, not a unique intended interpretation, as folk semantics would have led one to expect, but a family of interpretations that conform to community usage. It is too early to judge how well the program will succeed.

Field (1972) assumes that the mechanisms by which words get their referents will be simple and uniform, but that paper is, as we noted earlier, a product of wishful thinking. It is a fundamental tenet of the correspondence conception that the truth conditions of sentences are effects of the activities of the community of speakers, which means, presumably, that for each sentence there is a causal explanation of why that sentence has the truth conditions it has. As Stephen Leeds (1995) has emphasized, there is no good reason to suppose that these explanations will fit together to form a satisfying pattern. It may well be that the mechanisms by which sentences get their truth conditions are too variegated to afford any interesting generalizations. The correspondence conception will regard any significant etymological regularities as a welcome surprise.

I want to reiterate a point emphasized by Field, that what is at issue in discussions of the two concepts is more than just rivalry over which concept gets to wear the honorific title "truth." One cannot, without proper signage, use the word "truth" in a single context for both concepts, without causing confusion, but it doesn't much matter which concept gets to

bear the name. Field's preferred usage is to use "truth" for the thin notion and "determinate truth" for its weightier cousin.[15] That is the usage I'll follow here, although one might instead have used "truth" for the correspondence notion and introduced a new word for the disquotational usage.[16] The distinction is important; the name is not.

4. BLANKET ENDORSEMENT

In its broad theoretical role, correspondence truth isn't all that different from other concepts from the social sciences. The situation is different for disquotational truth, for reasons adduced in Field (1972). The simplest version of disquotationalism simply takes the (T)-sentences as axioms. A more sophisticated version defines truth in terms of reference and satisfaction, Tarski-style, starting with reference and satisfaction conditions for simple terms that it gets by simple enumeration. Either way, a key theoretical term is being introduced by listing its instances, without any explanation of what the items on the list have in common. Good scientific hygiene demands better. As Aristotle teaches us (*Metaphysics* A), a scientific theory should not only tell us what things there are but explain why things are as they are, and merely giving a list doesn't explain why true sentences are true.

A proper scientific theory should carve nature at the joints, in Plato's vivid phrase,[17] and when a key term is introduced merely by giving a list, the fear arises that things are being grouped together that don't have anything significant in common. Disquotational truth theory confirms this fear, since it tells us that the sentences that are true but not determinately true share no distinguishing feature. Even God with all His angels cannot partition the indeterminate sentences into "true" and "false."

If we are to accept disquotational truth as legitimate, we have to allow it a distinctive role, with different expectations from what we expect of theoretical terms generally. What permits such indulgence is to regard the (T)-sentences as having the same epistemic status as stipulative definitions. Each of the

(T)-sentences serves as a "partial definition," in Tarski's (1935, p. 264) phrase, true in virtue of the (disquotational) meaning of the word "true." Stipulative definitions don't have to correspond to anything in nature; we may define a term any way we like. If the definition doesn't carve nature at the joints and the defined term isn't projectable, the *definitum* is not likely to be useful, but the definition is perfectly legitimate nonetheless. If the definition consists of nothing more than a list of apparently unrelated items, the defined term is not going to have a helpful explanatory role, but the definition itself is permissible because a definition doesn't need to explain anything.

Not all definitions are stipulative. When Socrates asked, "What is justice?" he wasn't asking for a stipulation, and he certainly wasn't going to accept a list as an answer. He sought a so-called *real definition* of a concept already in use, a definition that gives the essence of the concept defined. Tarski took pains to avoid claiming anything more audacious about his proposed definition of truth than that it was "materially adequate,"[18] but he also insisted[19] that he was explicating a familiar notion, not introducing a new one. Inasmuch as truth is a term already in use, how can the disquotationalist claim that her theory of truth has the epistemic status of a stipulative definition? Isn't it rather closer to a philosophical analysis, thus vulnerable to Socratic refutation?

The answer is that the notion the disquotationalist is characterizing isn't the familiar notion. Under the pressure of trying simultaneously to serve as a causally explanatory notion and to provide the (T)-sentences, the colloquial concept has snapped in two, and the disquotationalist is fashioning a new concept out of the remains. A new concept means a fresh start. At the end of the day, we can ask whether then new notion is a legitimate successor to the original, but, as I indicated above, I don't think that's a very important question.

We may formulate an explicit, stipulative definition any way we like. However, the (T)-sentences don't literally take the form of an explicit definition; that is, taking the (T)-sentences as axiomatic doesn't provide us with a biconditional of the form,

x is true if and only if _____,

with no semantic terms in the blank. Oughtn't we worry that the disquotationalist is cheating, bullying us into accepting a substantive metaphysical thesis by insisting that the thesis has the "epistemic status" of a definition?

The disquotationalist has a one-word answer: conservativeness. We can be as creative as we like in postulating explicit definitions because explicit definitions are not creative. An explicit definition's only theoretical impact is to introduce a new word. Our success in making an explicit definition only depends on our intention to use a word in a particular way. It doesn't depend on the way things are with respect to the part of the world described by the original language. Technically, there are two notions of conservativeness in current usage, distinguished by Craig and Vaught (1958). A sentence or set of sentences $\triangle$ that introduces a new term is conservative over a background theory Γ *in the semantic sense* if every model of Γ can be expanded to a model of $\triangle$. $\triangle$ is conservative over Γ *in the proof theoretic sense* if every sentence of the original language that you can derive from $\Gamma \cup \triangle$ you can derive from Γ alone. The Completeness Theorem tells us that if $\triangle$ is conservative over the background theory in the semantic sense, than it's conservative in the proof-theoretic sense, but the converse does not hold. An explicit definition is conservative in both senses.

The theory of truth consisting of the (T)-sentences is conservative in both senses over any background syntactic theory that proves of any two different sentences that they are distinct. Thus the theory is as innocuous as an explicit definition. (This assumes that the word "true" doesn't appear in the background theory; the liar paradox is too big a can of worms to open here.)

An appeal to conservativeness guarantees that taking the (T)-sentences as axiomatic is methodologically permissible, but it leave us all the more puzzled why disquotational truth is of any use. What's the purpose of constructing a theory whose only effect is to define a term by lumping together things that don't have anything in common? The standard answer, due Quine (1986, pp. 10–13), treats the word "true" as a device for

generalization, in the same general line of work as universal quantification. We can express the grim lesson we learn from "Tom is mortal," "Dick is mortal," "Harry is mortal," and so on, by the generalization, "All men are mortal," and in the same way, we can proclaim the wisdom of "Tom is mortal or Tom is not mortal," "Snow is white or snow is not white," "Clare is poor or Clare is not poor," and so on, by declaring, "Every sentence of the form 'φ or not φ' is true." The strategy is better than the example. The example doesn't succeed in showing why the notion of truth is useful, since we were prepared to assent to statements of the form "φ or not φ" even before we were informed they were true.

The search for better examples has centered on statements like the following:

Everything the Pope says *ex cathedra* is true,

which has the effect of asserting infinitely many sentences of the form

If the Pope says that φ *ex cathedra*, then φ.

The "that φ" should give us pause. If we only allowed ourselves "purely disquotational" truth, which limits the application of the notion of truth to sentences of one's own current language, all we would be get would be instances of the schema

If the Pope says "φ" *ex cathedra,* then φ,

all of which are vacuously true, since the Pope seldom speaks English, and he never speaks English *ex cathedra*. To get the version with the "that" clauses, we require an interlinguistic notion of sameness of meaning, so that we have

If *ex cathedra* the Pope says something that means the same as "φ" means in my language, then φ.

One worries that perhaps the substantive word-to-world connections that characterize the correspondence theory are required to get the notion of sameness of meaning, so that the deflationary aspects of the disquotational theory are lost in translation. But let me set that worry aside.

The general principle that, if something can be recognized as true, then it is determinately true, holds across the board, for God as well as humans. If Clare is a genuine borderline case of "poor," then the reason you and I don't know whether Clare is poor isn't our ignorance. There is no fact of the matter there to be known, so that even God doesn't know. The reason for papal infallibility, so the story goes, is that God miraculously intervenes to make sure that the *ex cathedra* pronouncements of the pontiff are true. But God can only do this by making sure that the *ex cathedra* pronouncements are determinately true. Thus a sharper version of the doctrine is available

If the Pope says that φ *ex cathedra,* then φ is determinately true.

In other words, the thesis, "Everything the Pope says *ex cathedra* is true," holds even if truth is understood in the correspondence sense.

There is nothing special about the Pope in this. Whenever we are in a position to say that everything, or nearly everything, a person says of a particular topic is true, with "true" understood in the disquotational sense, we are in a position to say the same thing with "true"understood in the correspondence sense. Moreover, the formulation in terms of correspondence truth is better, not only because it is sharper, but because it is amenable to explanation. Papal infallibility is a supernatural phenomenon, so our usual expectations about explanations of regularities do not apply. In more ordinary circumstances, when we suppose that everything, or nearly everything, a person tells us on a particular topic is true, it's because we think the person is rigorous in her reasoning and honest and circumspect in her speech. But if one were trying to do semantics solely in disquotational terms, without acknowledging that true sentences have any significant properties in common, one would be hard put to explain why honesty, rigor, and circumspection are marks of truth.

So it goes with other common uses of the notion of truth. Things we want to say using the disquotational conception can be said as well or better using the correspondence notion. An

example Field mentions (1994, p. 120), is that we need the notion of truth to express the metaphysical realist thesis that there are true sentences that we will never have reason to believe. But for there to be true sentences that we will never have reason to believe, it is enough, on a disquotational reading of "true" (given classical logic[20]), that there be sentences about which we'll never have a reasoned belief one way or another. But merely to say that there are sentences about which we shall never have a reasoned opinion is surely not enough to commit a person to metaphysical realism. A realist believes that there are statements that are true *in the correspondence sense of "true"* without our having reason to believe them.

Again (1994, pp. 120f), one needs the notion to truth to formulate the norm of asserting truths and avoiding asserting untruths. But for this purpose, the two notions of truth serve equally well.

One needs the notion of truth (see 1994, p. 121) to repudiate at theory ("Some of the consequences of this theory are false") or merely to express doubt about it ("Perhaps some of the consequences of the theory are not true"). For this purpose, the correspondence notion will work as well as the disquotational.

Without a doubt, one can contrive examples – knights and knaves puzzles, things like that[21] – that rely on its being disquotational truth, rather than correspondence truth that is being employed. Moreover, disquotational truth is valuable because it's the bird in the hand. We have, at present, only the faintest sketch of what a correspondence theory of truth would look like. Lacking a theory of meaning fixation, what we can say now about correspondence truth serves merely to mark the place in conceptual space that a future theory is intended to occupy. Disquotational truth, by contrast, is well understood, and the two notions are nearly coextensive, so we use the notion as a placeholder for the notion of correspondence truth that will eventually emerge. However, we had hoped that disquotational truth could play a role more estimable then merely a theoretical stopgap or a trick for formulating brainteasers.

5. TRUTH AND PROOF

The place I propose to look for examples in which the disquotational notion of truth proves its worth is pure mathematics. The notion of truth functions as a powerful tool for mathematical discovery.[22] Consider the Gödel sentence for Peano arithmetic (PA). Nearly everyone who's thought about it and who's willing to accept arithmetic at all accepts the Gödel sentence,[23] but why? It's easy to see that the Gödel sentence follows from CON(PA), the statement that PA is consistent, but why should we believe that PA is consistent? One answer is to appeal to Gentzen's (1936) argument, but people accept CON(PA) without familiarity with Gentzen's proof. The simplest answer is that we regard the axioms as true, and the consequences of a true theory are consistent. We cannot formalize this argument within the language of arithmetic, because we have to go outside the language of arithmetic to define truth,[24] but once we have a standard semantic theory, we can prove that every consequence of a true theory is true, and hence that the true sentences are consistent, by a straightforward induction on the length of derivations. How do we convince ourselves that the axioms are true? I am not fretting here about the skeptical worry, "How do we know that there are numbers?" but rather wondering, in as much as there are infinitely many axioms, how to we get from acceptance of the axioms individually to the general thesis that all the axioms are true? One answer could be that we see the axioms are true by reflecting on our mathematical practice, but a more direct answer is possible. We can use the truth predicate to consolidate the infinitely many instances of the induction axiom schema into a single induction in the metalanguage, thus proving infinitely many axioms at one fell swoop. In short, we prove the Gödel sentence for PA by introducing a truth predicate for the language of arithmetic, then allowing the truth predicate to appear within instances of the induction axiom schema.

If we understand arithmetical language and we are willing to allow that there are such things as natural numbers, then we shall accept the principle of mathematical induction. If we

accept the principle of mathematical induction, we shall accept all the arithmetical sentences obtained by substituting an arithmetical formula into the induction axiom schema,

$$(R0 \wedge (\forall x)(Nx \rightarrow (Rx \rightarrow R(x+1)))) \rightarrow (\forall x)(Nx \rightarrow Rx)$$

and prefixing universal quantifiers to bind the variables, but these instances of the induction axiom schema don't exhaust what we know when we accept the principle of mathematical induction. The principle of mathematical induction allows us not only to accept instances of the induction axiom schema within the language of arithmetic, but instances of the schema formulated within whatever larger language we may get from the language of arithmetic by adding new predicates. In particular, adjoining a truth predicate to the language of arithmetic and permitting it to appear within induction axioms permits us to prove the Gödel sentence, CON(PA), and other useful theorems besides.

I made this argument in an earlier article, McGee (1997), and Field immediately saw its weak point:

> McGee seems to say that *G* [the Gödel sentence] is provable within schematic PA: all we have to do is add a truth predicate, and use inductions on it. But this is false: what is true (and what he says in his careful statement of the result) is that we get more powerful results if we add a truth predicate, use inductions on it, *and add a certain compositional theory of truth*. But of course, adding a compositional theory of truth is going beyond schematic arithmetic.[25]

If we don't include the compositional theory of truth, if we introduce the notion of truth simply by taking the (T)-sentences as axiomatic, then allowing the truth predicate to appear with induction axioms won't enable us to prove the Gödel sentence. In fact, we won't be able to prove any arithmetical sentences we couldn't prove before. This is shown by Ketland (1999). So what justifies a disquotationalist is accepting the compositional theory of truth?

Again, a one-word answer: conservativeness. The following characterization of truth, its *positive inductive definition*,[26] is

conservative, in both the semantic and the proof-theoretic senses, over whatever theory lies in the background:[27]

$(\forall x)(Tr(x) \leftrightarrow$
$[Sent(x) \wedge$
$[(x \text{ has the form } \tau = \rho \wedge Den(\tau) = Den(\rho))$
$\vee (x \text{ has the form } \sim \tau = \rho \wedge Den(\tau) \neq Den(\rho))$
$\vee (x \text{ has the form } \tau < \rho \wedge Den(\tau) < Den(\rho))$
$\vee (x \text{ has the form } \sim \tau < \rho \wedge Den(\tau) \not< Den(\rho))$
$\vee (x \text{ has the form } (\varphi \vee \psi) \wedge (Tr(\ulcorner \varphi \urcorner) \vee Tr(\ulcorner \psi \urcorner)))$
$\vee (x \text{ has the form } \sim(\varphi \vee \psi) \wedge (Tr(\ulcorner \sim \varphi \urcorner) \wedge Tr(\ulcorner \sim \psi(\tau) \urcorner)))$
$\vee (x \text{ has the form } (\exists v)\varphi(v) \wedge (\exists \text{ closed term } \tau) Tr(\ulcorner \varphi(\tau) \urcorner))$
$\vee (x \text{ has the form } \sim(\exists v)\varphi(v) \wedge (\forall \text{ closed term } \tau) Tr(\ulcorner \sim \varphi(\tau) \urcorner))$
$\vee (x \text{ has the form } \sim \sim \varphi \wedge Tr(\ulcorner \varphi \urcorner))]]).$

Here "*Den*" abbreviates the primitive recursive function that takes a closed term to the number it denotes.[28] Thus the very same feature that justifies us in treating the (T)-sentences as true by definition permits us to regard the positive inductive characterization as definitional.

The positive inductive definition differs from the usual compositional theory in that it doesn't supply the principle of bivalence, the thesis that every sentence is either true or false, but not both. But once we have the positive inductive definition, we can prove bivalence by an induction on the complexity of sentences, and once we have bivalence we can prove the Gödel sentence, CON(PA), and the rest.[29]

What distinguishes a disquotational from a correspondence conception of truth is that the former regards semantic theory as true by definition, whereas the latter seeks a semantic theory that is true in virtue of the activities of a community of speakers. In order to plausibly regard the semantic theory as definitional, it has to be conservative, but within that constraint a number of different theories are possible. One of them, the theory consisting of the naked (T)-sentences, is useless mathematically, but another, the positive inductive definition, is enormously helpful.

With the right system of axioms, a disquotational conception of truth can be a valuable mathematical tool. Is that

really a reason to favor a disquotational theory, or can we get the same benefits using a correspondence theory instead? The two conceptions diverge in their treatment of semantically defective sentences, but for the language of arithmetic, presumably, there are no semantically defective sentences. Setting the skeptical worries of Field (1998) aside, the intended models of the language of arithmetic are determined uniquely up to isomorphism, and so every arithmetical sentence is either determinately true or determinately false. Even so, the results we obtain by employing a correspondence conception are less than satisfactory. On a correspondence account, bivalence is a contingent fact about the linguistic practices of our community, and it is only in virtue of this fact that determinate truth – what the correspondence theorist calls simply "truth" – and (disquotational) truth are coextensive. But we surely don't want our acceptance of the Gödel sentence to depend on results from sociolinguistics. The correspondence-theoretic version of the proof of the Gödel sentence is less than satisfactory.

There is a different way to establish CON(PA) that doesn't bring in any philosophy at all, namely, to derive it from the axioms of set theory. This strikes me as not fully satisfactory, both because it is a little unnatural to drag set theory into the picture and because, to both philosophers and mathematicians, set theory has seemed less secure than number theory. In any case, the same issues will come up when we ask about CON(ZFC).

For the language of arithmetic, we were able to give a direct compositional characterization of truth, because every individual in the intended domain of the model is named by some closed term. We can't do the same thing for set theory. For set theory we have to either enlarge the language by adding a name for every set or else give a compositional theory of satisfaction and then define truth in terms of satisfaction. There's no way the former approach can yield a correspondence theory of truth, since it's not possible for the activities of a community of speakers to affix a denotation to uncountably many individual constants. Expanding the lan-

guage to include a name for every pure set only makes sense if the names are regarded as mathematical abstractions, not dependent on the practices of the community of set theorists. The definition of truth in terms of satisfaction might perhaps be regarded as a correspondence theory, but for the reasons given in Field (1972), it will be a dreadfully inadequate theory unless it's supplemented with an account of why "$\in$" refers as it does. Either account can be regarded as a legitimate disquotational theory, for the same reason as before: conservativeness. The compositional theory of truth for the expanded language and the compositional theory of satisfaction for the original language can both be formulated as first-order positive inductive definitions, and first-order positive inductive definitions are always conservative.

The proofs of CON(ZFC) and of the Gödel sentence for ZFC proceeds just like the corresponding proofs for PA, with the replacement axiom schema taking the place of the induction axiom schema.[30] The difference is that for the language of set theory there is a worry about semantically defective sentences even for people who aren't inclined toward skepticism. In (1997), I proposed that it was possible to give a categorical characterization of the universe of pure sets, but this is by no means the predominant view. Perhaps the most widely held view, dating back to Zermelo (1930), is that there are many "universes" of set theory. Whenever you take an intended model of set theory and clip off the construction at some (uncountable strong) inaccessible level, you get another intended model. For any two nonisomorphic universes, one is isomorphic to an initial segment of the other, got by cutting off the cumulative hierarchy at some inaccessible, but there is no master universe that contains all the others. If this picture is right, then, presumably, a sentence will count as determinately true – true in the correspondence sense – if it is true in every universe.[31]

The argument that ZFC can be seen to be consistent because all the theorems of ZFC are true goes through on either conception of truth. But consider instead the following conditional

There is an inaccessible → CON(ZFC + "There is an inaccessible").

The conditional counts as true on either conception of truth, because "CON(ZFC + 'There is an inaccessible')" is an arithmetical statement, and arithmetical statements don't change from one universe to another. But can we prove it, just using the correspondence notion? The set of determinately true sentences is consistent, and all the consequences of ZFC are determinately true. But "There is an inaccessible" isn't determinately true, and so the proof falls apart. The consistency of the set of determinate truths doesn't get us the consistency of ZFC + "There is an inaccessible." To prove the consistency of ZFC + "There is an inaccessible," we need disquotational truth; correspondence truth can't do the job.

In summary, I want to agree with the contention that there are two legitimate concepts of truth. We need correspondence truth to solve the Maiasaur Problem, and we need disquotational truth as a tool for discovering new mathematical truths.[32]

NOTES

[1] The (T)-sentences are sentences that follow the paradigm, "'Snow is white' is true if and only if snow is white."

[2] I intend the phrase "linguistic practices of speakers" to be understood liberally, so that it encompasses speakers' mental states as well as their dispositions to verbal behavior, since what a speaker means by an expression depends, in part, on her state of mind when she uses it. Also, talk about the effects of linguistic practices needs to make allowance for the fact, dramatically illustrated by Putnam's (1975) "Twin Earth" example, that semantic values depend not only on the things speakers do and think but on the situations in which they do them and think them.

[3] Perhaps one should say instead that it is the quotation name "'seal'" that is ambiguous, standing for either of two words that are spelled and pronounced alike but mean different things. I don't have a firm opinion about this, although I do want to insist that "mass," as nineteenth-century physicists used it, was a single word, as is "true" as it is ordinarily used today.

[4] This important example was brought to the philosophical community's attention by Field (1973).

[5] This is pointed out in the postscript to Field (1972), which is in Field (2001, pp. 27–29).

[6] Cf. Tarski (1935, pp. 153 and 263).

[7] See Dummett (1975).

[8] Lewis (1991, pp. 57–59); see also Burgess (2004).

[9] See Gentzen (1969).

[10] There are limits to how far one can push this sort to thing, illustrated dramatically by Prior (1961). My own take on the issue is developed in (2000, 2001).

[11] This thesis was proposed by van Fraassen (1966), who was anticipated by Mehlberg (1958), and further developed, among others, by Lewis (1970), Fine (1975), Kamp (1975), Field (1994a), and McGee and McLaughlin (1995).

[12] There has been some perplexity in the literature, going back to Fine (1975, p. 143), about whether van Fraassen's semantics sanction conditional proof and *reductio ad absurdum*. If we assume φ and derive ψ, can we discharge to conclude $(\varphi \rightarrow \psi)$? If our "derivation" of ψ from φ only included inferences permitted by classical logic, then the answer is assuredly "Yes," but some authors, notably Williamson (2004), have argued that a thorough commitment to classical logic should allow the intermediate derivation to include any inference with the property that the conclusion is true in all acceptable models if the premises are true in all acceptable models, whether or not that inference is sanctioned by classical logic. If the derivation of ψ is only held to this more relaxed standard, then we can't be safe in concluding $(\varphi \rightarrow \psi)$. In McGee and McLaughlin (2004), we propose an intermediate standard: the "good" inferences – the ones that can be legitimately employed even within indirect proofs – should all have the property that the conclusion is true in every acceptable model in which each of the premises are true. This standard validates conditional proof and *reductio ad absurdum*.

[13] http://www.rom.on.ca/palaeo/maiasaur.

[14] Unger (1980, 1979); see also Wheeler (1979) and Lewis (1993). The thesis that the conclusion of Unger's argument should be that reference is inscrutable, rather than (as Unger himself would have it) that there are no clouds or people is developed in McGee and Mclaughlin (2000) and McGee (2004, 2005).

[15] An apparent advantage of this way of talking is that the operator "determinately" can be applied to other predicates, so that we can say "determinately poor" and "determinately red." Because of the inscrutability of reference, this device is not as useful as one might have hoped. If "Clare" had a uniquely picked out referent, we could say that "Clare is poor" is determinately true if and only if Clare is determinately poor; but "Clare" doesn't have a uniquely determined referent.

[16] In McGee and McLaughlin (1995), we suggest "pluth," an elision of "pleonastic truth."
[17] *Phaedrus* 265e.
[18] Tarski (1935, p. 152, and 1944, pp. 69–71).
[19] Tarski (1944, p. 69).
[20] From the (T)-sentences, we are able to derive all instances of the excluded-middle schema ($\varphi \vee \sim\varphi$), which gives us full classical logic, within either the strong or the weak 3-valued logic of Kleene (1952), section 54.
[21] These are Raymond Smullyan's (1978, 1980, 1982) delightful puzzles about an island occupied by knights (who always tell the truth), and knaves (who always lie).
[22] This proposal is developed more thoroughly in McGee (to appear).
[23] Field himself has his doubts; see Field (1998).
[24] See Tarski (1935), pp. 274–276.
[25] Emphasis in original. P. 355 of the postscript to Field (1998), pp. 351–360 of Field (2001).
[26] For an illuminating general treatment, see Moschovakis (1974).
[27] Thus the theory, in virtue of its form as a first-order positive inductive definition, is conservative by the strictest possible standard, conservative over pure logic, whereas the theory consisting of the (T)-sentences is only conservative over a background theory that includes a modicum of syntax; see Halbach (2001). We do, however, require a bit of syntactic theory to derive the (T)-sentences from the positive inductive definition.
[28] The logical primitives of the language are disjunction, negation, existential quantification, and identity.
[29] If we take the usual compositional theory, including bivalence, as axiomatic, we get an extension of PA (which includes the syntax by way of Gödel numbering) that is conservative in the proof-theoretic sense but not in the semantic sense. This is a deep result of Kotlarksi, Krajewski, and Lachlan (1981) and Lachlan (1981); see also Kaye (1991, chap. 15). I give my reasons for thinking that proof-theoretic conservativeness isn't good enough in (to appear).
[30] I assume that the axiomatization is given in such a way that the separation and foundation axioms are derivable from replacement axioms.
[31] An attractive alternative treats a sentence as true if it is true in all sufficiently large universes. The point I'm making here, that a correspondence theory of truth can't get the job done by itself, still goes through on this alternative conception, but we have to change the example; see my (to appear).
[32] A version of Chapter 3 was read to a IAP lecture at MIT and to the philosophy colloquium at the University of California Irvine. A version of Chapter 5 was read to the conference on Self-reference organized by the Danish Network for Philosophical Logic and its Applications in Copen-

hagen and to the Arché group at St. Andrews. I am grateful for the help I received.

REFERENCES

Burgess, J. (2004): 'Mathematics and *Bleak House*, *Philosophia Mathematica* 12, 18–36.

Craig, W. and Vaught. R.L. (1958): 'Finite Axiomatizability Using Additional Predicates', *Journal of Symbolic Logic* 23, 289–308.

Dummett, M.A.E. (1975): 'Wang's Paradox', *Synthese* 30, 301–324. Reprinted in Keefe and Smith (1996), pp. 99–118.

Field, H. (1972): 'Tarski's Theory of Truth', *Journal of Philosophy* 69, 347–375. Reprinted in Field (2001), pp. 3–26.

Field, (1973): 'Theory Change and Indeterminacy of Reference', *Journal of Philosophy* 70, 462–481. Reprinted in Field (2001), pp. 177–193.

Field, H. (1974): 'Quine and the Correspondence Theory', *Philosophical Review* 83, 200–228. Reprinted in Field (2001), pp. 199–218.

Field, H. (1977): 'Logic, Meaning, and Conceptual Role', *Journal of Philosophy* 74, 379–409.

Field, H. (1986): 'The Deflationary Conception of Truth', in G. MacDonald and C. Wright (eds.), *Fact, Science and Value* (pp. 55–117) Oxford: Blackwell.

Field, H. (1994): 'Deflationist Views of Meaning and Content', *Mind* 103, 249–285. Reprinted in Field (2001), pp. 104–140. Page references are to the reprint.

Field, H. (1994a): 'Disquotational Truth and Factually Defective Discourse', *Philosophical Review* 103, 405–452. Reprinted in Field (2001), pp. 222–258.

Field, H. (1998): 'Which Undecidable Mathematical Sentences Have Determinate Truth Values?', in H. G. Dales and G. Oliveri (eds), *Truth in Mathematics* (pp. 291–310), Oxford: Oxford University Press. Reprinted in Field (2001), pp. 332–350. Page references are to the reprint.

Field, H. (2001): *Truth and the Absence of Fact,* Oxford: Clarendon Press.

Fine, K. (1975): 'Vagueness, Truth, and Logic', *Synthese* 30, 265–300. Reprinted in Keefe and Smith (1996), pp. 119–150. Page references are to the reprint.

Gentzen, G. (1936): 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* 112, 493–565. English translation by M. H. Szabo in Gentzen (1969), pp. 132–213.

Gentzen, G. (1969): *Collected Papers,* Amsterdam and London: North-Holland.

Halbach, V. (2001): 'How Innocent is Deflationism?', *Synthese* 126, 167–194.

Kamp, J.A.W. (1975): 'Two Theories about Adjectives', in E.L. Keenan (ed.), *Formal Semantics of Natural Language* (pp. 123–155), Cambridge: Cambridge University Press.
Kaplan, D. (1989): 'Demonstratives', in J. Almog (ed.), *Themes from Kaplan*, (pp. 481–564), Oxford and New York: Oxford University Press.
Kaye, R. (1991): *Models of Peano Arithmetic,* Oxford: Clarendon Press.
Keefe, R. and Smith, P. (1996): *Vagueness: A Reader,* Cambridge, Mass., and London: MIT Press.
Ketland, J. (1999): 'Deflationism and Tarski's Paradise', *Mind* 108, 69–94.
Kleene, S.C. (1952): *Introduction to Metamathematics,* New York: American Elsevier.
Kotlarski, H., Krajewski, S. and Lachlan, A.H. (1981): 'Construction of Satisfaction Classes for Nonstandard Models', *Canadian Mathematical Bulletin* 24, 283–293.
Lachlan, A.H. (1981): 'Full Satisfaction Classes and Recursive Saturation', *Canadian Mathematical Bulletin* 24, 295–297.
Leeds, S. (1995): 'Truth, Correspondence, and Success', *Philosophical Studies* 79, 1–36.
Lewis, D.K. (1970): 'General Semantics', *Synthese* 22, 18–65. Reprinted in Lewis (1983), pp. 189–229.
Lewis, D.K. (1983): *Philosophical Papers*, vol. 1, New York and Oxford: Oxford University Press.
Lewis, D.K. (1991): *Parts of Classes,* Oxford and Cambridge, Mass.: Blackwell.
Lewis, D.K. (1993): 'Many but Almost One', in J. Bacon, K. Campbell and L. Reinhardt (eds.), *Ontology, Causality, and Mind*, (pp. 23–38), New York: Cambridge University Press.
Martinich, A.P., (ed.) (2000): *Philosophy of Language* 4th edn., New York and Oxford: Oxford University Press.
McGee, V. (1997): 'How We Learn Mathematical Language', *Philosophical Review* 106, 34–68.
McGee, V. (2000): 'Everything', in G. Sher and R. Tieszen (eds.), *Between Logic and Intuition*, (pp. 54–78), New York and Cambridge: Cambridge University Press.
McGee, V. (2001): 'Truth by Default', *Philosophia Mathematica* 9, 5–20.
McGee, V. 'In Praise of the Free Lunch', to appear in V.F. Hendricks, S.A. Pedersen and T. Bollander (eds.), *Self-Reference* (Stanford: CSLI).
McGee, V. (2005): 'Inscrutability and its Discontents', *Noûs* 39, 397–425.
McGee, V. (2004): 'The Many Lives of Ebeneezer Wilkes Smith', in G. Link (ed.), *One Hundred Years of Russell's Paradox*, (pp. 611–624), Berlin: Walter de Gruyter.
McGee, V. and McLaughlin, B.P. (1995): 'Distinctions Without a Difference', *Southern Journal of Philosophy* 33 supplement (Spindel Conference volume for 1994), 203–252.

McGee, V. and McLaughlin, B.P. (1998): Review of Williamson (1994). *Linguistics and Philosophy* 21, 221–235.
McGee, V. and McLaughlin, B.P. (2000): 'The Lessons of the Many', *Philosophical Topics* 28, 128–151.
McGee, V. and McLaughlin, B.P. (2004): 'Logical Commitment: A Reply to Williamson', *Linguistics and Philosophy* 27, 123–136.
Mehlberg, H. (1958): *The Reach of Science,* Toronto: University of Toronto Press. Except reprinted in Keefe and Smith (1996), pp. 85–88.
Moschovakis, Y.N. (1974): *Elementary Induction on Abstract Structures,* Amsterdam: North-Holland.
Prior, A. (1961): 'The Runaway Inference Ticket', *Analysis* 21, 38–39. Reprinted in Strawson (1967), pp. 129–131.
Putnam, H. (1975): 'The Meaning of 'Meaning', in K. Gunderson (ed.), *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science*, vol 7 (pp. 131–193), Minneapolis: University of Minnesota Press. Reprinted in Putnam (1975a), pp. 215–271.
Putnam, H. (1975a): *Mind, Language, and Reality. Philosophical Papers,* vol. 2. Cambridge: Cambridge University Press.
Quine, W.V. (1960): *Word and Object,* Cambridge, Mass.: MIT Press.
Quine, W.V. (1968): 'Ontological Relativity', *Journal of Philosophy* 65, 185–212. Reprinted in Quine (1969), pp. 26–68.
Quine, W.V. (1969): *Ontological Relativity and Other Essays,* New York: Columbia University Press.
Quine, W.V. (1970): *Philosophy of Logic* 2nd edn., Cambridge, Mass., and London: Harvard University Press.
Russell, B. (1923): 'Vagueness', *Australasian Journal of Philosophy and Psychology* 1, 84–92. Reprinted in Keefe and Smith (1991), pp. 61–68. Page references are to the reprint.
Smullyan, R.M. (1978): *What is the Name of this Book?* Englewood Cliffs, New Jersey: Prentiss-Hall.
Smullyan, R.M. (1980): *This Book Needs No Title*. Englewood Cliffs, New Jersey: Prentiss-Hall.
Smullyan, R.M. (1982): *The Lady or the Tiger?* New York: Knopf.
Soames, S. (1984): 'What is a Theory of Truth?', *Journal of Philosophy* 81, 411–429.
Stalnaker, R. (1984): *Inquiry,* Cambridge, Mass.: MIT Press.
Strawson, P.F. (ed.) (1967): *Philosophical Logic,* Oxford: Oxford University Press.
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. Page references are to the translation.
Tarski, A. (1936): 'Über den Begriff der logischen Folgerung', *Actes du Congrès International de Philosophie Scientifique* 7, 1–11. English translation by J. H. Woodger in Tarski (1983), pp. 409–420.

Tarski, A. (1944): 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research* 4, 341–375. Reprinted in Martinich (2000), pp. 69–91. Page references are to the reprint.

Tarski, A. (1983): *Logic, Semantics, Metamathematics*, 2nd edn., Indianapolis: Hackett.

Unger, P. (1979): 'I Do Not Exist', in G. MacDonald (ed.), *Perception and Identity*, (pp. 235–251), Ithaca, N.Y.: Cornell University Press.

Unger, P. (1980): 'The Problem of the Many', *Midwest Studies in Philosophy* 5, 411–467.

Van Fraassen, B.C. (1966): 'Singular Terms, Truth-value Gaps, and Free Logic', *Journal of Philosophy* 63, 481–495.

Wheeler, S.C. (1979): 'On That Which Is Not', *Synthese* 41, 155–173.

Williamson, T. (1994): *Vagueness,* London: Routledge.

Williamson, T. (2004): 'Reply to McGee and McLaughlin', *Linguistics and Philosophy* 27, 93–111.

Zermelo, E. (1930): 'Über Grenzzahlen und Mengenbereiche', *Fundamenta Mathematicae* 16, 29–37.

Massachusetts Institute of Technology
Department of Linguistics and Philosophy
77 Massachusetts Avenue
Building 32-D931
Cambridge MA 02139
USA
E-mail: vmcgee@mit.edu