

Chapter 4 Deflationist Views of Meaning and Content

1. Two views of meaning and content

I see the philosophy of language, and the part of philosophy of mind concerned with intentional states like believing and desiring and intending and the like, as pretty much bifurcated into two traditions. The traditions differ over the role that the notion of truth conditions plays in the theory of meaning and in the theory of the content of intentional states.

One of the traditions, whose early advocates include Frege, Russell, the early Wittgenstein, and Ramsey, has it that truth conditions play an extremely central role in semantics and the theory of mind; a theory of meaning or content is, at least in large part, a theory of truth conditions. A strong *prima facie* reason for the attractiveness of this position is that the way we standardly ascribe meanings and contents is via "that" clauses, and the ascription of "that" clauses is in effect the ascription of truth conditions: to describe an utterance as meaning that snow is white, or a belief state as a state of believing that snow is white, is in effect to say that the utterance or belief state has the truth conditions that snow is white. Since "that" clauses and hence truth conditions play such a central role in our *ascriptions* of meaning and content, it would seem as if they ought to play a central role in the *theory* of meaning and content. However, it isn't easy to say precisely what this central role is that truth conditions allegedly play; that I think is a main motivation for the alternative tradition.

As a crude paradigm of the other tradition, consider the verifiability theory of meaning. Here the main concept is not truth conditions but verification conditions. The verification conditions of a type of utterance might be given by the class of sensory stimulations that would or should lead to the acceptance of an utterance in that class. Notice that if the verification conditions of a type of utterance are given in this way, then they are given without a "that" clause; and indeed, there is no immediately clear way in which to obtain a "that" clause from them. It is because of this that an advocate of the Frege-Russell-Ramsey-*Tractatus* tradition is likely to regard a verification theory of meaning as leaving out what is central to semantics: a theory of meaning according to which you can fully describe the "meaning" of an utterance without saying that it means *that snow is white* (or that something else) has left out the central element of meaning. That central element is truth conditions.

Some advocates of the Frege-Russell tradition might prefer to describe what has been left out not as truth conditions but as *propositional content*. There is a way of understanding this that makes it unobjectionable, but it could be misleading. For a verificationist is likely to simply deny the charge that he has left out propositional content (or even, the charge that he has demoted it from a central place in his theory). A proposition, he can say, is simply a class of verification conditions; for an utterance or belief state to express the proposition is for it to have verification conditions in that class. So propositions in the verificationist's sense needn't be described by "that" clauses: the most direct way to specify a proposition in the verificationist's sense is directly in terms of verification conditions. No one can deny that propositions *so understood* play a central role in the verificationist's theory. An advocate of the Frege-Russell tradition is likely to say that these "propositions" aren't what *she* means by propositions, and that is doubtless so; that is because propositions as she conceives them must *encapsulate truth conditions*. (They may encapsulate more than truth conditions, but they encapsulate at least this much.) So it seems to me that to describe the difference between the two traditions in terms of truth conditions rather than in terms of propositional content brings to the forefront what is really central.

I do not mean to suggest that a verificationist is precluded from speaking of truth or truth conditions; he can do so in either of two ways. One way, which seems to me a very bad idea, is to introduce some epistemic notion of truth conditions: that is, to define truth in terms of verification (e.g. "would be verified in the long run"), so that truth conditions are derivatively defined in terms of verification conditions. I won't discuss this line here. The other way is to make use of what has been called a *deflationary* conception of truth. I take this to have several variants, but the variant I will primarily focus on is called *pure disquotational* truth. As a rough heuristic, we could say that for a person to call an utterance true in this pure disquotational sense is to say that it is true-as-he-understands-it. (This is *not* intended to provide a *definition* of pure disquotational truth in terms of some other notion of truth plus a notion of understanding; it is intended only as a heuristic, to motivate the features of pure disquotational truth I now proceed to enumerate.) As the heuristic suggests, a person can meaningfully apply "true" in the pure disquotational sense only to utterances that he has some understanding of; and for such an utterance *u*, the claim that *u* is true (true-as-he-understands-it) is cognitively equivalent (for the person) to *u* itself (as he understands it).¹ A qualification is needed, since the claim that *u* is true involves an existential commitment to the utterance *u*, whereas *u* itself doesn't; this

keeps the two from being fully cognitively equivalent. The qualified version is that the claim that *u* is true is cognitively equivalent to *u* relative to the existence of the utterance *u*; just as "Thatcher is such that she is self-identical and snow is white" is cognitively equivalent to "Snow is white" relative to the existence of Thatcher. (To say that A is cognitively equivalent to B relative to C means that the conjunction of A and C is cognitively equivalent to the conjunction of B and C; so that as long as C is presupposed we can treat A and B as equivalent.) Having made this qualification, I will generally leave it tacit, and simply say that in the purely disquotational sense of "true", the claim that *u* is true (where *u* is an utterance I understand) is cognitively equivalent for me to *u* itself (as I understand it).²

There are both oddities and attractive features to this as a reading of "true"; I will discuss these later on, starting in Section 5. But odd or not, the cognitive equivalence of the claim that *u* is disquotationally true to *u* itself provides a way to understand disquotational truth independent of any nondisquotational concept of truth or truth conditions (and independent of any concept of proposition). That is: if I understand "Snow is white", and if I also understand a notion of disquotational truth as explained above, then I will understand "'Snow is white' is true", since it will be equivalent to "Snow is white". This will hold on any view of understanding, even a crude verificationist view according to which understanding a sentence is simply a matter of having verification conditions for it: for the verification conditions of "'Snow is white' is true" will be precisely those of "Snow is white". Certainly there is no need to presuppose that understanding an utterance involves correlating a proposition or truth condition with it, as propositions and truth conditions are understood in the Frege-Russell tradition. Because of this, there can be little doubt that the notion of pure disquotational truth is a notion which a verificationist is entitled to and which other opponents of the Frege-Russell tradition who have a more sophisticated alternative than verificationism are entitled to as well. (Adherents of the Frege-Russell tradition are entitled to it too, though not all want it.)

And using this notion of truth, even a crude verificationist can grant the legitimacy of talk of truth conditions of his own utterances. For the cognitive equivalence of "'Snow is white' is true" and "Snow is white" will lead to the (more or less indefeasible) acceptance of the biconditional "'Snow is white' is true iff snow is white"; and a natural way to put this (more or less indefeasible) acceptance is to say "'Snow is white' has the truth conditions that snow is white". A pure disquotational notion of truth gives rise to a purely disquotational way of talking about truth conditions.

An opponent of the Frege-Russell tradition can try to use this purely disquotational way of talking about truth conditions, or some variant or extension of it, to allow the legitimacy of talk of the truth conditions of utterances and mental states without giving truth conditions the central role that they play in the Frege-Russell tradition. For instance, if one of our mental states can be described as an attitude of believing or accepting the sentence "Snow is white", then it can be described both as a state of believing that snow is white and as a state with the truth conditions that snow is white. The connection of "that" clauses to truth conditions noted in the second paragraph thus remains, but the initial impression that it leads inevitably to the Frege-Russell view is at least lessened. It may not be obvious that all legitimate talk of the truth conditions of mental states can be handled in this way or by an extension of it, but that is what the opponent of the Frege-Russell tradition contends. (This may involve disallowing as illegitimate certain uses of truth conditional talk that an advocate of the Frege-Russell view would allow.)³ Such a view might naturally be called a *deflationist view of meaning and content*; or more accurately, a deflationist view of *meaning that* and *having the content that*, or a deflationist view about *the role of truth conditions in meaning and content*. The first of the three labels (which I have used in the title for brevity) could be misleading, since some versions of the view are in a sense quite undeflationist about meanings: for instance, the crude verificationism I described is not in the least deflationist about verification conditions; and it identifies them with meanings. What it is deflationist about is truth conditions; or, the locutions "means *that*" and "has the content *that*". From now on I'll simply use the label "deflationist". So that my labels not be thought as prejudging the merits of the views, I'll call the Frege-Russell tradition "inflationist".

The division between the inflationist and deflationist positions is in some ways the most fundamental division within the theory of content and meaning (though as with many fundamental divisions in philosophy, the line between the two views is not absolutely sharp). I myself strongly feel the attractions of each position, though I have come to favor deflationism. In this paper I will formulate a fairly radical version of deflationism,⁴ say a few things to try to motivate it, and try to defend it against some obvious objections. I will also mention some deeper lines of objection that I cannot deal with here; these seem to me the places where the battle between deflationism and inflationism must ultimately be fought.

2. More on the deflationist/inflationist distinction

The version of deflationism I outlined in my opening remarks was built around a crude verificationism. It

hardly needs arguing that such a version of deflationism isn't satisfactory. But the main idea behind deflationism doesn't require verificationism, it requires only that what plays a central role in meaning and content not include truth conditions (or relations to propositions, where propositions are conceived as encapsulating truth conditions).

I must admit at the start that the question of whether truth conditions (or propositions encapsulating truth conditions) play a central role in meaning and content is not a very clear one, for three reasons. First, the notions of meaning and content aren't clear as they stand. Some ways of clarifying them involve legislating truth conditions into or out of meaning and content, which would make the issue of deflationism totally uninteresting. My way of resolving this unclarity will be to interpret "meaning" and "content" as broadly as possible short of explicitly legislating truth conditions into meaning or content. Second, the idea of a *central* role isn't initially clear either. My goal is to clarify this as I go along, by making explicit the limited role that I think the deflationist can give to truth conditions, and identifying kinds of role that deflationism cannot allow truth conditions to have. Unfortunately I cannot complete the job in this paper: to keep the paper to a manageable size I have had to leave to another occasion discussion of the crucial subject of the role of truth conditions in psychological explanation. [But see Section 7 of the Postscript.]

There is also a third source of unclarity to the question, one which I think makes a complete sharpening impossible. If deflationism is to be at all interesting, it must claim not merely that what plays a central role in meaning and content not include truth conditions *under that description*, but that it not include *anything that could plausibly constitute a reduction of* truth conditions to other more physicalistic terms. And it is well known that there is no sharp line between reduction and elimination, so this crucial tenet of deflationism is not an altogether precise one. In other words, a theory of content and meaning might end up not employing the notion of truth conditions directly (in a central role), but employing (in that role) certain physicalistic relations that *could* be regarded as reducing the relation "S has the truth conditions p", but could *only* be so regarded on a rather loose conception of reduction. Such a theory of content and meaning would occupy the borderline between a deflationist and inflationist theory: it would be rather a matter of taste which way we describe it. I'll have a bit more to say about this possibility later.

I remarked earlier that a deflationist need not be a verificationist. What elements instead of, or in

addition to, verification conditions does he have available for inclusion in meaning or content? One element that can certainly be included in content is conceptual or computational role: role in a (perhaps idealized) computational psychology that describes how the agent's beliefs, desires, etc. evolve over time (partly in response to sensory stimulations).⁵ The conceptual role of a belief state includes its verification conditions, but includes much more besides. (It includes a wider variety of evidential relations;⁶ it includes the conceptual consequences of having the belief; etc.) But conceptual role isn't enough: it is both "internalist" and "individualist", and a plausible deflationism is going to have to give to content both "externalist" and "social" aspects.

In describing these further aspects of content, it will help if I make what I hope is a harmless assumption. The assumption is that we can speak of a language-user as believing and desiring sentences of his or her own language or at least, as believing and desiring internal analogs of them in which some or all of the ambiguities may have been removed and that this relation can be made sense of without a prior notion of content for the belief or desire states or of meaning for the sentences. In other words, this relation is one that a deflationist can perfectly well appeal to. There is a good bit that needs to be said about just what this assumption comes to, but let me simply say that I think that when it is unpacked it is pretty unobjectionable.⁷

Now let's get back to the aspects of content that go beyond conceptual role. The most obvious "externalist" element that a deflationist can put into content is indication relations. It is a fact about me that I am a pretty good barometer of whether there is rain falling on my head at that moment: when there is rain falling on my head, I tend to believe "There is rain falling on my head"; conversely, when I do believe this sentence, usually there is rain falling on my head. This is simply a correlation, there to be observed; and a deflationist is as free to take note of it as is anyone else, and as free as anyone else to deem it an ingredient of what he calls content. The correlation of people's belief states to the world outside presumably extends past the directly observed; the beliefs of astronomers in sentences of form "The location of Halley's comet at time t will be x " generally correlate pretty well with the location of Halley's comet at time t , and in certain circumstances the beliefs of physicists in sentences of form "A particle has recently tunneled through the potential barrier" will correlate with whether a particle has indeed done so.

These observations might make one think that a deflationist is bound to recognize (a non-

disquotational version of) the relation "S has truth conditions p", in fact if not in name, since these indication relations constitute the truth-conditions relation. But this overlooks the fact that the project of giving anything close to a believable reduction of talk of truth conditions to talk of indication relations is at best a gleam in the eye of some theorists. A way to see the point is to notice that there are plenty of examples where the indication relations don't reflect what we would intuitively regard as truth conditions. Consider the ancient Greek whose judgements (that we would translate as) >Zeus is throwing thunderbolts= are reliably correlated with the presence of lightning in the vicinity. That reliable correlation is a key fact about the use of the sentence, and is part of its *content* in the deflationist's sense, but it is not a matter of the truth conditions of his utterance as normally understood. (It would be possible to maintain that the Greek's judgements had as their truth conditions that lightning is present; I wouldn't want to say that that is exactly wrong, since I doubt that there is a clear matter of right and wrong here, but it certainly fits ill with the fact that there were systematic inferential connections between this sentence and others containing >Zeus=, such as >Let's build an idol so Zeus will stop throwing thunderbolts=.) In many other cases, the facts about indication relations are clear but a decision about truth conditions seems rather arbitrary. The judgement >The sun is rising= today may have as its truth conditions that the angle from the horizon to the sun is small, positive and increasing, but was this so at earlier times when people believed that its motion relative to the earth was absolute motion? I have no idea how to decide, but from a deflationist point of view it doesn't matter: the indication relations are there anyway, and they are what is important.

I have given examples from earlier theories since we obviously can't recognize any such serious errors that may infect our own observational claims, but it is not beyond imagination that even some of our own observational reports reliably indicate something at variance with their own truth conditions. When we move beyond observation reports to more theoretical sentences, this is still more likely.

There are of course less systematic indication relations than the ones I have considered. Maybe I systematically exaggerate, so that my believing a sentence of the form "It is n feet high" is strongly correlated with the object before me being f(n) feet high, where f(x) starts dropping off rapidly from x after about 6 feet or so. Or worse, maybe my beliefs of the form "In Bosnia, p" don't stand in any interesting correlations with the actual facts about what's happening in Bosnia, but just reflect what appears in the newspapers I read. The deflationist can recognize *these* facts about indication too, and attribute explanatory importance to them. The deflationist can also use the disquotation schema to

formulate the distinction between these cases and cases like my reliability about whether rain is falling on my head: in the latter case my belief state reliably indicates its own truth conditions (as given by the disquotation schema), in the former it indicates something else.⁸ What he can't do, it seems to me, is say that this distinction is of much explanatory importance: for that would give truth conditions (rather than just indication relations in general) a central role in the theory of mind; and the claim that truth conditions have a central role in the theory of mind is the defining characteristic of inflationism.⁹ But that isn't to say that he can't differentiate the exaggeration and Bosnia cases from the rain falling on my head case on other grounds, e.g. of how systematic the indication relations are and how well they survive under the addition of evidence. And here it is worth noting that on these scores the Zeus example and the rising sun example (and analogous examples of as-yet-undetected errors in our own observation reports) come out more like the rain falling on my head example than like the exaggeration or Bosnia examples: they are rather systematic indication relations not dislodged by any small amount of additional evidence. It is these factors, rather than the factor of whether what is indicated is truth conditions, that are explanatorily important.

Anyway, the deflationist can include indication relations in content; this is enough to make content externalist, but not enough to make it social. However, since we are regarding many of our most important beliefs as in effect attitudes toward sentences in our language, the means to make content social is at hand. That is, it seems reasonable to decree that the content of these belief states is to be influenced by the meaning these sentences have for us; that meaning is of course partly influenced by the content of our intentional states, but since language is social it will be influenced by the contents of the states of other members of our community as well. In particular—and removing the appearance of circularity—we can understand "content" in such a way that the conceptual roles and indication relations of other people's states of believing a certain sentence are counted as relevant to the content of my state of believing that sentence;¹⁰ such a way of understanding "content" is not only possible, it is natural when one thinks of the ways that *which sentences I believe* influence and are influenced by *which sentences others believe*. Again, construing content as social in this way is quite compatible with deflationism, that is, with keeping truth conditions (and hence "that" clauses) out of the fundamental characterization of content.

3. The semantics of logical operators

"Deflationism" and "inflationism" are broad labels, each encompassing many different views. I think it would help to make the contrast between the deflationist and inflationist traditions clearer if I focussed on several problems in the theory of content, and showed how they looked from each of the two broad perspectives.

The first problem I will discuss is the problem of saying what determines the "referent" of a logical operator like "or" or "not" or "all"; or to put it less Fregeanly, the problem of saying what determines the contributions that such a logical operator makes to the truth conditions of sentences that contain it.

I think it is clear in broad outline what an inflationist's best strategy is for answering this question. The answer will have two stages. The first stage involves spelling out the "conceptual role" that "or" has (for a given person, or for the linguistic community generally): in this case, it would largely be a matter of its role in deductive inference and perhaps inductive and practical inference as well. For instance, among the relevant facts about my use of the word "or" is that I tend to accept inferences to a disjunction "A or B" from either disjunct, and to the negation of the disjunction from the negations of both disjuncts, but not to the disjuncts from the disjunction; and when I do occasionally slip and infer according to the latter rule, I can be brought to correct myself. Also relevant perhaps is that I tend to assign to a disjunction a degree of belief higher than the degrees of belief of each disjunct but no higher than the sum of the degrees of belief of the disjuncts. (Indeed my degree of belief in a disjunction "A or B" tends to be about equal to the sum of my degrees of belief in the disjuncts minus my degree of belief in A times my conditional degree of belief in B given A.) And so on.

But what does the word's having this conceptual role have to do with its standing for the usual truth function (that is, with its contributing to truth conditions according to the usual truth table)? Perhaps the conceptual role of "or" determines that it stands for that truth function *in this sense*: any word in any linguistic community that had that conceptual role would stand for that truth function. But that doesn't seem very helpful, for it doesn't tell you why that conceptual role should be associated with that truth function.

There is, though, a rather natural story to try to tell. The idea is that if we assign truth conditions to our sentences (and hence our belief states) according to this rule of truth for "or" (and the standard rules of truth for "not"), it will make our deductive inferences involving "or" truth-preserving, and presumably will make our inductive inferences involving "or" highly reliable;¹¹ whereas if we used any other truth table, the deductive and inductive inferences would come out totally unreliable. Reliability considerations seem to give just the sort of link between conceptual role and reference that we want in this case.

There are of course some holes here. One hole is that though it is true that no alternative *truth function* makes our inferences involving "or" at all reliable, this just shows that reliability considerations make the usual truth table for "or" more satisfactory *than alternative truth functions*; but maybe there is an alternative to a truth functional account that would make us equally reliable? A second hole is that in explaining the contribution of "or" to truth conditions, we can't legitimately assume the contribution that "not" makes to truth conditions unless that can be independently established; so unless it can be independently established (which is doubtful), or unless we can avoid mentioning it in our account of what determines the truth table for "or", then we really need an account of how reliability considerations give us the rules of truth for "or" and "not" together. (Using another operator like "if...then" or "implies" instead of "not" would give rise to similar problems.) This deepens the first hole: for now we need to compare the package containing the usual truth functional accounts of "or" and "not" to alternative packages in which both truth rules are varied, and where there is no constraint that either truth rule be truth functional.

Even if the second hole can be avoided, the problem with the first hole is greater for more complicated operators, like quantifiers: for instance, in the case of "some" our usual deductive inferences turn out truth preserving if we suppose that its contribution to truth conditions is that of the unrestricted objectual existential quantifier, but making it a restricted quantifier or substitutional quantifier of a certain sort would also make the inferences truth preserving, and (depending on the details of the restriction) might make us at least equally reliable. Even with the truth rules for the truth functional connectives and "implies" unchanged.

So what determines that our word "or" or "not" or "some" makes the contribution to truth conditions that we assume it makes? An inflationist has three choices. Either she can find some sort of naturalistic facts that could be cited in answer to this question (facts involving reliability considerations might work in the end, despite the questions above); or she can say that there is simply no fact of the matter, that we have a surprising example of referential indeterminacy here; or she can say that it is a non-naturalistic fact about us that our word "some" contributes to truth conditions according to the rules of the unrestricted objectual quantifier and that "not" and "or" contribute according to the usual truth tables. But this third option strikes me as rather repellent; the first may be hard to carry out; and the second may not be altogether free of philosophical difficulties.

The same sort of problem arises in the case of the predicate "is identical to" (and the numerical quantifiers such as "there are exactly two" that are definable from it). The axioms that we accept as governing this predicate determine (given a reliability assumption) that it stands for a congruence relation, that is, an equivalence relation for which substitutivity holds; but this is not enough to determine that it stands for genuine identity, that is, for the relation that everything bears to itself and to nothing else. It is especially not enough if the extensions of the other predicates in the language are up for grabs: as Quine has famously pointed out, it isn't so easy to say what facts about us make "identical to" stand for identity and "rabbit" for rabbits, rather than "rabbit" for rabbit stages and "is identical to" for the relation of being stages of the same object. And even if we can find facts about usage that would rule out this example as violating a plausible reliability requirement, it is a tall order to find facts about usage which would rule out *all* choices of nonstandard congruence relations. Again, an inflationist is apparently faced with the three options of finding such facts, or accepting it as simply a brute fact that our word stands for genuine identity, or accepting a surprising level of referential indeterminacy in our basic logical vocabulary. This is a choice that not everyone would be happy to have to make.

A main motivation for deflationism is that it apparently avoids having to make this sort of choice.¹² According to the simplest version of deflationism that relying on the pure disquotational truth predicate mentioned earlier it is an entirely trivial matter to explain why one's own word "or" obeys the truth table it does: this follows from truth functional logic together with the logic of the disquotational truth predicate, with no mention of any facts at all about our usage of "or". Recall that if "true" is used purely disquotationally, then "p is true" is cognitively equivalent to "p", for any sentence "p" that we understand. As a result, we get that each instance of the "disquotation schema"

(T) "p" is true if and only if p

holds of conceptual necessity, that is, by virtue of the cognitive equivalence of the left and right hand sides. But for any sentences q and r, instances of the following schemas are also instances of (T), and hence conceptually necessary:

"q or r" is true if and only if q or r

"q" is true if and only if q

"r" is true if and only if r.

And now you can use the latter two to substitute into the right hand side of the first, so as to get the conceptual necessity of each instance of the following:

(TF) "q or r" is true if and only if "q" is true or "r" is true.

The conceptual necessity of the instances of (TF) really isn't enough to justify the claim that "or" obeys the usual truth table: we need the conceptual necessity of the generalization

(TFG) For all sentences S_1 and S_2 of our language, $\text{Å}S_1 \text{ or } S_2\text{Å}$ is true iff S_1 is true or S_2 is true.

But it is easy to get that if we start not from (T) but from some generalized form of (T). One such generalized form of (T) employs a universal substitutional quantifier:

(TG) $(\forall p[\text{"p" is true if and only if p}])$.¹³

(I assume a theory of substitutional quantification that avoids the semantic paradoxes.) There are also ways to get the generalization without substitutional quantifiers. One alternative is to incorporate schematic letters for sentences into the language, reasoning with them as with variables; and then to employ two rules of inference governing them: (i) a rule that allows replacement of all instances of a schematic letter by a sentence; (ii) a rule that allows inference of $\exists x(\text{Sentence}(x) \in A(x))$ from the schema $A(\text{"p"})$, where $A(\text{"p"})$ is a schema in which all occurrences of the schematic letter p are surrounded by quotes. Such a formalism corresponds to a very weak fragment of a substitutional quantifier language, and is probably preferable to using the full substitutional quantifier. In such a formalism, (T) and (TF) themselves are part of the language, rather than merely having instances that are part of the language; and from (T) we can derive (TF) and thence (by rule (ii) and a bit of syntax) (TFG).

Note that on any of these versions of the pure deflationary account, facts about the meaning of our word "or" need not be explicitly referred to in explaining why our word "or" obeys the usual truth table. Of course, we do need to use logical reasoning, and in particular, the deductive inferential rules that are in fact associated with the word "or". But we needn't mention in the explanation that these rules are associated with "or"; in that sense, the meaning of "or" plays no role in the explanation.

Does this mean that a deflationist can't make sense of the idea that "or" obeys the truth table it

does *because of* its conceptual role? Not really: the fact that *the deductive inferential rules for "or"* are used in the explanation of "or" having the truth table it has is enough to give one quite natural sense to the claim that it obeys that truth table *because* it obeys those deductive inferential rules.

However, the sense of "because" here does not support counterfactuals. That is, because *the fact that these rules are associated with "or"* plays no role in the explanation of "or" having this truth table, we don't get the conclusion that if different rules had been associated with "or", its contributions to truth conditions would have been different. (At least, we don't get this conclusion on the most straightforward reading of the counterfactual.) Indeed, it is clear that this counterfactual (on its straightforward construal anyway) is unacceptable if "truth conditions" is understood in a purely disquotational sense. (This may seem an objection to the purely disquotational notion of truth, but later on I will argue that it is not. I will also consider whether a deflationist can make sense of a modified notion of truth conditions according to which such a counterfactual would be straightforwardly true; if so, the conceptual role of "or" might have to be explicitly referred to in explaining why "or"'s truth conditions in this modified sense of truth conditions are those of the usual truth table.)

What I've said here about the explanation of the truth-theoretic properties of the logical connectives goes for other words as well. Consider "rabbit": an inflationist presumably thinks that the set or property that my term "rabbit" stands for is determined from the facts about this word's conceptual role for me, together with its conceptual role for other members of my community, together with the facts about what my believing various "rabbit" sentences tends to be correlated with, together with the same sort of facts for other members of my community, and so on. This raises the question of precisely how it is determined; and it seems to me that if inflationism is to be believable then the inflationist needs to have some story to tell here. Of course, Quine's well-known difficulty about rabbits v. rabbit-stages would be solved if we solved the problem for identity: reliability with respect to sentences like "For all x and y, if x and y are nearby rabbits then $x=y$ " ties the extension of "rabbit" to that of "identical" in such a way as to handle such individuation problems. But there are further problems: for instance, what makes "rabbit" stand for the rabbits rather than the rabbits-or-realistic-imitations? I don't say that the inflationist can't tell a reasonable story about this, only that there is a story to be told, and perhaps there is room for skepticism about the possibility of telling it adequately. If so, that provides a motivation for deflationism. For the deflationist view is that there is nothing here to explain: it is simply part of the logic of "refers" (or "is true of") that "rabbit" refers to (is true of) rabbits and to nothing else.

The deflationary view of truth and reference as I've presented it here applies only to words and sentences that we understand. This may well seem worrisome, and it is important to ask both whether it *should* seem worrisome and whether it could be avoided. I'll return to these matters in later sections, but first I want to further explore the contrast between the deflationist and inflationist viewpoints.

4. Inverting the theory of reference

One qualm that one might reasonably have about the deflationist perspective is that a lot of work that has gone into the theory of reference in recent years seems to be onto something, and it seems at first hard to explain just what it could be onto if truth conditions play no central role in the theory of meaning. After all, if truth conditions play no central role, reference can hardly play a central role: whatever importance reference has surely derives from its contribution to truth conditions.

Let's consider first Kripke's observation (Kripke 1972) that description theories of the reference of our proper names¹⁴ are incorrect. If truth conditions play no central role in meaning, and truth is fully explained by the disquotation schema (and of value only as a logical device, in a manner soon to be explained), then the same is true of reference: for the reference of singular terms, the relevant schema is

(R) If b exists then "b" refers to b and nothing else; if b doesn't exist then "b" doesn't refer to anything.

But if this tells us everything we need to know about the reference for our own words, what could Kripke's critique of description theories be telling us?

We need a concrete example before us. Consider Kripke's case against a version of the description theory according to which the referent of our term "Godel" is determined by the associated description "prover of the incompleteness theorem": the referent of the name (if it has one) is to be whatever uniquely fits the description. Kripke's case against this view is a thought experiment, in which we discover the following fact:

(F) The incompleteness theorem was proved by a man baptized "Schmidt" and who never called himself anything other than "Schmidt"; a certain person who called himself "Godel" and got a job under that name at the Institute for Advanced Study stole the proof from him.

In this situation, Kripke asks, who is it natural to say we have been referring to when we used the name "Godel", the guy who called himself "Schmidt" and proved the incompleteness theorem, or the guy who called himself "Godel"? Nearly everyone says the latter, so the description theory (or this version of it) is wrong. Surely there is something important in this critique, and if the deflationist can't make sense of it then something is wrong with deflationism.

I think that the deflationist can make sense of Kripke's observations. On the deflationist viewpoint, though, *the observations aren't at the most basic level about reference but about our inferential practice*. That is, what Kripke's example really shows is that we would regard the claim (F) as *grounds for inferring* "Godel didn't prove the incompleteness theorem" rather than as grounds for inferring "Godel was baptized as 'Schmidt' and never called himself 'Godel'". Reference is just disquotational. It does come in indirectly: from (F) I can indirectly infer

"Godel" doesn't refer to the guy that proved the incompleteness theorem.

But that isn't because of a causal theory of reference over a description theory, but only because I can infer

Godel isn't the guy that proved the incompleteness theorem,
and then "semantically ascend".

This does seem to me a fairly plausible account of what the Kripke point shows. We see that at least in this case, the deflationist picture leads to a reasonably plausible inversion of standard views, and tends to demote the importance of theories of reference and truth conditions from accounts of language use and cognitive functioning.

But how about the positive view of reference that emerges from the writings of Kripke, Putnam, etc.? The positive view is that a typical name of mine refers to what it does because of a causal network of beliefs and utterances involving that name: in the simplest cases, some people acquired beliefs involving a name as a result of direct causal interaction with the thing named; these beliefs led to their using the name in utterances, which led other people to have beliefs involving the name; these in turn passed beliefs on to others, and so forth. And it is because my uses of (say) "Hume" stand in a causal network of roughly this sort, a network whose dominant causal source is Hume, that my uses of the name refer to Hume. The previous discussion shows how a deflationist can partially capture the force of

this: he can say that it is just part of our inferential procedure to regard claims of roughly the form "The dominant causal source of our beliefs involving 'b' is b" as pretty much indefeasible. But this doesn't seem to fully capture the importance of the causal theory of reference. It has seemed to many that the causal network emanating from Hume to my uses of the name "Hume" explains what is otherwise mysterious, namely, how my name could be *about* Hume. Obviously a deflationist can't say this: the whole point of deflationism is that the only explanation we need of why my word is about Hume is given by the disquotation schema.

Nonetheless, the deflationist can agree that this causal network story is explanatory of something: what it explains is the otherwise mysterious correlation between a knowledgeable person's beliefs involving the name "Hume" and the facts about Hume. You probably believe quite a few sentences that involve the name "Hume", and a large proportion of them are probably disquotationally true: that is, the conditional probability that Hume was Φ given that you believe "Hume was Φ " is quite high. Surely this correlation between your "Hume" beliefs and the Hume-facts cries out for explanation.¹⁵ And the general lines of the explanation are clearly suggested in the Kripke-Putnam account: you acquired your "Hume" beliefs largely through interactions with others, who in turn acquired theirs from others, and so on until we reach believers with a fairly immediate causal access to Hume or his writings or whatever. Moreover, the causal network has multiple independent chains, and contains historical experts who have investigated these independent chains systematically, so the chance of large errors surviving isn't that high. The role of "experts" that figures heavily in Putnam's account of reference thus also has its analog when focus is put instead on explaining our reliability. Indeed, when the focus is put on reliability it is obvious *why* the reliance on "experts" should have such a central place.

Earlier I pointed out that the reliability of my beliefs under the disquotation schema is simply an objective fact about me, stateable without semantic terms, which a deflationist can hardly be debarred from taking note of in his account of meaning. Here I am expanding the point, to observe that he can hardly be debarred from wanting an explanation of it; and the explanation is bound to involve some of the ingredients that inflationists tend to put into their theories of reference.

This of course reraises the possibility mentioned earlier, that by the time the deflationist is finished explaining this and similar facts, he will have reconstructed the inflationist's relation "S has the truth conditions p", in fact if not in name. My guess is that this will turn out not to be the case, but to

assert that conclusion with confidence requires a more thorough investigation than I will be able to undertake in this paper. All I really hope to motivate here is that we should be "methodological deflationists": that is, we should start out assuming deflationism as a working hypothesis; we should adhere to it unless and until we find ourselves reconstructing what amounts to the inflationist's relation "S has the truth conditions p". So methodological deflationism is simply a methodological policy, which if pursued could lead to the discovery that deflationism in the original sense ("metaphysical deflationism") is workable or could lead to the discovery that inflationism is inevitable. It could also turn out that we end up constructing something that *might or might not* be regarded as the inflationist's relation "S has the truth conditions p"; in that case, the line between inflationism and metaphysical deflationism will turn out to have blurred.¹⁶

5. More on disquotational truth

It is time now to be a bit more precise about what exactly deflationism involves. In order to do this, it will help to ask what a deflationist should say about why we need a truth predicate. If truth conditions aren't central to meaning, why not drop talk of truth altogether? As is well known, the deflationist's answer is that the word "true" has an important logical role: it allows us to formulate certain infinite conjunctions and disjunctions that can't be formulated otherwise.¹⁷

There are some very mundane examples of this, for instance, where we remember that someone said something false yesterday but can't remember what it was. What we are remembering is equivalent to the infinite disjunction of all sentences of form "She said 'p', but not-p".

A more important example of this which has been widely noted arises in discussions of realism. "Realism" has been used to mean many things, but one version of it is the view that there might be (and almost certainly are) sentences of our language that are true that we will never have reason to believe, in contrast to "anti-realist" doctrines that identify truth with long-run justifiability or whatever. To assert realism in this sense one needs a notion of truth. But the reason for this is purely logical: that is, if only finitely many sentences could be formulated in our language, we could put the realist doctrine without use of a predicate of truth: we could say:

It might be the case that either the number of brontosaurus that ever lived is precisely 75,278

but we will never have reason to believe that; or the amount that Michael Jackson spent on underwear in his lifetime is exactly \$1,078,852.72 but we will never have reason to believe that; or

where in place of the "... " goes similar clauses for every sentence of the language. It is because we can't complete the disjunction that we need a notion of truth. Or perhaps I should say, because we can't *otherwise* complete the disjunction: for the claim that it might be the case that some sentence of our language is true but that we will never have reason to believe it can be viewed as simply a way of formulating the disjunction.¹⁸

Another example of "true" as a device of infinite conjunction and disjunction is the desire to utter only true sentences or to have only true beliefs: what we desire is the infinite conjunction of all claims of the form "I utter 'p' only if p" or "I believe 'p' only if p". It is sometimes claimed that a deflationist cannot grant that there is any "substantial norm" of assertion beyond warranted assertibility.¹⁹ This seems to me a serious mistake: any sane deflationist will hold that truth and warranted assertibility (even long run warranted assertibility) can and do diverge: as the previous paragraph should make clear, their divergence is a consequence of the truth schema together with quite uncontroversial facts. Consequently, a norm of asserting the truth is a norm that goes beyond warranted assertibility. But there is no difficulty in desiring that all one's beliefs be disquotationally true; and not only can each of us desire such things, there can be a general practice of badgering others into having such desires. Isn't this enough for there being a "norm" of asserting and believing the truth? Admittedly, this account of norms in terms of badgering is a bit crude, but I see no reason to think that on a more sophisticated account of what a norm is, norms of striving for the truth won't be just as available to the deflationist as they are on the crude account.

I'll give a fourth example of the utility of a device of infinite conjunction and disjunction, since it clearly brings out some points I want to stress. Consider some theory about the physical world, formulated with finitely many separate axioms and finitely many axiom schemas (each schema having infinitely many axioms as instances). For an example of such a theory, one can take a typical first order version of the Euclidean theory of space (which is not finitely axiomatizable). Suppose that one rejects this theory without knowing which specific part of it to reject; or alternatively, suppose that one accepts it but regards it as contingent. In the first case one will put the rejection by saying "Not every axiom of

this theory is true"; in the second case, by saying "It might have been the case that not every axiom of the theory was true". But the intended purpose in the first case is of course to deny the infinite conjunction of the axioms, and in the second case to assert the *possibility* of the negation of this infinite conjunction.

My reason for focussing on this example is that it shows clearly the importance of what I earlier called pure disquotational truth. Pure disquotational truth involves two important features that certain other truth concepts lack. The first is that I can understand "Utterance u is true" only to the extent that I can understand utterance u; the second is that for me, the claim that utterance u is true in the pure disquotational sense is cognitively equivalent to u itself as I understand it. (The first pretty much follows from the second: if I can't understand u, and if an attribution of truth to u is cognitively equivalent to u, then I can't understand an attribution of truth to u.)²⁰ The second feature of the pure disquotational notion of truth means that this notion is of a *use-independent* property: to call "Snow is white" disquotationally true is simply to call snow white; hence it is *not* to attribute it a property that it wouldn't have had if I and other English speakers had used words differently.

In ordinary English we seem to use a truth predicate that does not share these two features. It is not entirely obvious, though, that what we say can't be understood indirectly in terms of disquotational truth. (More on this in Sections 8 and 9.) In any case, these two features of pure disquotational truth make it ideally suited to serve the logical need for a device of infinite conjunction and disjunction illustrated in the examples above, particularly the Euclidean geometry example. In the first place, the only sentences we ever literally conjoin or disjoin are sentences we understand, so it is clear that a notion that is inapplicable to utterances we don't understand will serve our needs for truth as a device of conjunction and disjunction. Second and more important, the use-independence of disquotational truth is *required* for the purposes just reviewed. For if "All sentences of type Q are true" is to serve as an infinite conjunction of all sentences of type Q, then we want it to entail each such sentence, and be entailed by all of them together. This would fail to be so unless "S is true" entailed and was entailed by S. But the only way that can be so is if "true" doesn't ascribe a use-dependent feature to S. Suppose for instance that Euclidean geometry is true, and that we try to express its contingency by saying that the axioms together might have been false. Surely what we wanted to say wasn't simply that speakers might have used their words in such a way that the axioms weren't true, it is that space itself might have differed so as to make the axioms *as we understand them* not true. A use-independent notion of

truth is precisely what we require.

Of course, there is another use-independent notion of truth besides disquotational truth that would be usable for these purposes: truth as applied to *propositions*, where propositions are construed as language-independent (rather than, say, as equivalence classes of utterances). But I think that pure disquotational truth is better for these purposes, for two reasons. The first and less important is ontological: it is better to avoid postulating strange entities unnecessarily. The more important reason is that unless one is very careful to limit the use made of the notion of expressing a proposition, the introduction of such a notion of proposition would beg the question in favor of inflationism.²¹ There may be good arguments for inflationism, but the need to express infinite conjunctions and disjunctions can be handled with the more minimal apparatus of purely disquotational truth, and it is best to use only the more minimal apparatus as long as the need of the more powerful apparatus is unargued.

I have been arguing that inflationists as well as deflationists need some use-independent notion of truth as a device of infinite conjunction and disjunction, and that such a notion is needed only for sentences that one understands. Pure disquotational truth serves the need admirably; and I think that a deflationist should take pure disquotational truth as the fundamental truth concept. This is not to say that it is the only notion of truth available to the deflationist: in Section 9 I will consider the possibility of introducing a modified kind of disquotational truth, which both applies to sentences in languages we can't understand and is not use-independent. But I have argued that a deflationist needs a notion of pure disquotational truth, and I think that he should regard it as the basic notion of truth in terms of which others are to be explained.

6. Disquotational truth versus Tarskian truth

In this Section I consider the relation between Tarskian truth definitions and disquotational truth. To facilitate the discussion I will pretend that English is a language without indexicals or demonstratives or ambiguous sentences, so that we can apply talk of truth to sentence types. (I will consider what the deflationist should say about indexicality, ambiguity and the like in Section 10.)

As I have explained disquotational truth, axiomatic status is given to some generalized version of the truth schema

(T) "p" is true if and only if p.

The generalized version might be the result of prefixing the schema with a universal substitutional quantifier; alternatively, we might prefer the weaker approach involving schematic variables, mentioned earlier. (Again, I assume sufficient restrictions on the schema to avoid semantic paradoxes.) Actually, if our language contains a modal operator we need something a bit stronger, as the Euclidean geometry example of the previous section indicates: the schema (T) should be replaced by:

(NT) \sim ["p" is true if and only if p].

But even when the language doesn't have a modal operator, the left hand side of (T) is to be understood as cognitively equivalent to the right hand side. (Any view, no matter how inflationist, accepts the instances of (T) as material biconditionals.)

Tarskian approaches are somewhat different. First I should remind the reader that though Tarski was interested in explicitly defined truth predicates, he also proved a severe limitation of such explicitly defined predicates: they are available only for fragments of one's language. If we explicitly define a truth predicate in a language, not only must we exclude its application to some or all sentences of that language that contain semantic terms, we must also exclude its application to a substantial body of other sentences of the language without semantic terms but with unrestricted quantifiers. And this of course limits the use of the truth predicate as a device of infinite conjunction and disjunction: for instance, we can't use it to form infinite conjunctions and disjunctions of sentences with unrestricted quantifiers, at least if arbitrarily large numbers of alternations of those quantifiers appear among the conjuncts or disjuncts. If we want to avoid such limitations on formulating infinite conjunctions and disjunctions, we should give up on the idea that our truth predicate must be explicitly defined.

Even if a Tarskian were to give up the insistence on *defining* truth, there would still be an important difference between the Tarskian approach and the full-fledged deflationary approach that takes a generalized version of the truth schema as an axiom. For central to what is usually thought of as the Tarskian approach (though perhaps it owes more to Davidson than to Tarski himself) is that truth²² is characterized (inductively if not explicitly) in terms of compositional structure.²³ This gives compositional principles of truth a much more central role than they have on the full-fledged deflationary account. Recall the deflationist story I gave about why the word "or" obeys the usual truth table: the principle

If S_1 and S_2 are sentences of our language then $\neg(S_1 \text{ or } S_2)$ is true iff S_1 is true or S_2 is true is built directly into the inductive characterization of truth on the Tarskian approach, whereas on the deflationary approach it is simply a consequence of the generalized truth schema (together with the principle $\forall S(S \text{ is a sentence} \supset \neg p(S='p'))$, on the substitutional quantifier version of the generalized schema). Something similar holds for other compositional principles, such as that a sentence consisting of a 1-place predicate and a referring name is true iff the predicate is true of what the name denotes: on a deflationist account this is simply a consequence of generalized disquotational schemas for "true", "true of" and "refers" (together with principles like the one just given parenthetically, together with some obvious syntactic principles and laws of concatenation and quotation marks).²⁴

Why does it matter whether such principles are built into an inductive characterization of truth or viewed as consequences of generalized disquotational schemas? One reason why it matters is that there is no guarantee that for all regions of discourse there will be such compositional principles that follow from the generalized disquotation schemas. The compositional principles for "or" and other truth functional operators simply fall out of the logical rules of inference that govern those operators (together with the disquotational principles for "true"); the same is basically true of other standard compositional principles (except that we sometimes need subsidiary semantic notions such as satisfaction, and disquotational principles governing them). But there is no obvious reason to think that in the case of all operators, e.g. non-extensional ones, there will be subsidiary semantic notions that yield compositional principles.

This shouldn't upset the full-fledged deflationist: for the full-fledged deflationist, such compositional principles have no particular interest in their own right; we can explain why we have them when we do, since they follow from the basic disquotational schemas whenever a substitutivity principle is part of the logic, but they are not of foundational importance. For the Tarskian they are of much more fundamental importance: they are needed in order to have a satisfactory notion of truth. This is one important way in which a Tarskian approach isn't *fully* deflationist.

A Tarskian approach may or may not be *partially* deflationist: whether it is depends on the status that is accorded the Tarskian "truth definitions". In particular, what (if anything) makes such a definition *correct* for a given population? For a view to count as in any serious sense deflationist, it

must say either that there is no sense at all to be made of speaking of one Tarski-predicate for a population as "correct", or that the only sense that can be made of it is based on stipulating that the *homophonic* truth predicate counts as correct *for us*.

To elaborate: it is easy to use Tarski-predicates in a theory that is in no way deflationist. One view of the significance of a Tarskian truth definition (whether inductive or explicit) is as partially characterizing an *abstract language*, that is, a language viewed as an abstract entity that exists whether or not anyone actually speaks it: the language is defined by the rules governing it, including Tarskian truth rules. (See Lewis 1975; Soames 1984; Stalnaker 1984.) So in addition to English there is an abstract language English*, with the same rules of grammar but different truth rules: e.g. in English* "or" might obey the truth table for conjunction and "rabbit" might be true of all and only the dinosaurs. Then questions about the truth conditions of a person's sentences are relocated as questions about which abstract language she speaks. An inflationist thinks that there are facts about a person's employment of her sentences by virtue of which it is one abstract language rather than another that she is speaking; such facts determine that Aristotle spoke in abstract Greek, and in the same way determine that we are speaking abstract English. The problem is to say what these facts are, and how they do the determining. But a deflationist thinks that a homophony condition guarantees that we are speaking English rather than English*, without the need of any such facts about our employment of the language.²⁵

There is of course a strong suggestion of the deflationist viewpoint in Tarski himself, stemming from his famous adequacy condition on truth definitions, "Convention T". (I should actually say, a strong suggestion of the *partial* deflationist view, because of the points earlier in this section.) My point is only that if you deemphasize these and focus on the truth definitions themselves, and think it is a straightforwardly factual question which truth definition is correct for a given population, then there is nothing in the least deflationist about the resulting view.

7. Objections to deflationism

There are many different sorts of worries that one might have about deflationism. Some seem to me to be based on misunderstandings. Under this heading I would include the commonly-voiced worry that deflationism cuts language off from the world. The worry is that if I simply accept "T-sentences" like

"There are gravitational waves" is true iff there are gravitational waves, then unless the sentence on the right hand side already has truth conditions of a more-than-disquotational sort, it has nothing to do with gravitational waves, and so the T-sentence doesn't really supply truth conditions to "There are gravitational waves". The response is that we do indeed need to establish a connection between the use of "There are gravitational waves" and gravitational waves, independent of the truth schema. But deflationism allows for this (at least in the version I have sketched): it admits that among the important facts about the use of "There are gravitational waves" are facts, stateable quite independent of disquotational truth, that relate the sentence to gravitational waves. For instance, the laws of physics are such that gravitational waves, if they exist, will cause pulses in a quadrupole antenna, and such pulses are one of the things that would increase our confidence in the sentence "There are gravitational waves". The objection tacitly assumed that the only kind of connection between gravitational waves and the use of "There are gravitational waves" (other than the derivative connection that arises simply from the use of "true" in accordance with the disquotation schema) comes in the non-disquotational assignment of truth conditions to that sentence; but this is simply false.

In contrast, many of the worries that one might have about deflationism raise deep issues. In my view the most serious worry about deflationism is that it can't make sense of the explanatory role of truth conditions: e.g. their role in explaining behavior, or their role in explaining the extent to which behavior is successful. Unfortunately it is a big job even to state the worry clearly, and a bigger job to answer it; I must save this for another occasion.²⁶ Another kind of worry about deflationism that is too big to discuss here is that it can't make the distinction between vague and non-vague discourse, or between discourse that is thoroughly factual and discourse that falls short of this in various ways. I discuss this worry in Chapter 8.

But there remain some less serious objections that I can discuss here; discussing them will help clarify what I am taking deflationism to involve. The objections to be discussed are

1. That deflationism can't handle the attribution of truth to utterances in other languages.
 2. That deflationism gives the wrong modal properties to "true".
 3. That deflationism can't handle ambiguous utterances, indexicals, demonstratives, and the like.
- and
4. That it can't make sense of how we learn from others.

I think all of these can be easily answered, but it is worth spelling out how this is to be done, for my treatment will separate my version of deflationism from certain other less radical versions. I will treat these objections in order in the final four Sections.

8. Applying "true" to sentences in other languages

I have emphasized that a person can meaningfully apply a pure disquotational truth predicate only to utterances that he understands. This raises the question of what we are doing when we talk of the truth or falsity of sentences in other languages.

There are at least three different approaches that a deflationist might take toward talk of truth for sentences in other languages. Perhaps the most obvious but to my mind the least satisfactory would be to use a notion of interlinguistic synonymy: regard "S is true" (where "S" stands for a foreign sentence) as equivalent to "S is synonymous with a sentence of ours that is true in the purely disquotational sense". This option brings the semantic notion of synonymy into the characterization of truth for foreign sentences. I will call truth as so defined "extended disquotational truth".

Some may object that extended disquotational truth is not a notion that a deflationist can allow himself, since it involves interlinguistic synonymy, which must be understood as involving sameness of truth conditions. I hesitate to say anything so strong; I say only that a deflationist who wants to employ a notion of interlinguistic synonymy faces a serious challenge, the challenge of showing how to make sense of it without relying on a prior notion of truth conditions. (Recall that the deflationist/inflationist contrast was explained in terms of whether truth conditions play a role in semantics and the theory of content.) But perhaps this challenge can be met. If for instance one were to adopt a crude verificationism, synonymy would be just interpersonal sameness of verification conditions assuming of course that that can be spelled out. On a more reasonable deflationist view, meaning will involve a variety of components: as discussed earlier, the meaning of a sentence of mine will depend on its inferential role for me, its inferential role for my fellow speakers, and its indication relations for me and my fellow speakers. So whether one of my sentences is synonymous with a sentence used by someone else (e.g., a foreign speaker) will depend on a variety of disparate facts. In that case there is no obvious way to get a useful and well-behaved synonymy relation.²⁷ Still, it isn't idiotic to suppose that there is an unobvious way

to do it; so it isn't initially clear that a deflationist can't admit a useful and well-behaved notion of extended disquotational truth.

But extended disquotational truth isn't the only option the deflationist has available for talking about truth conditions for foreign sentences. A second option is more flexible: simply define what it is for a foreign sentence to be *true relative to a correlation* of it to one of our sentences: a sentence is true relative to the correlation if the sentence of ours correlated with it is true in the purely disquotational sense. This approach differs from the previous in that no claim is made that the correlated sentences are synonymous. Of course, if we accept the notion of interpersonal synonymy, we can use it to get the effect of the first option: we can say that a foreign sentence S is true in the extended disquotational sense iff there is a correlation of S to a *synonymous* English sentence such that S is true relative to this correlation. But the second option can be applied by those like Quine who reject a notion of interpersonal synonymy, or by those who accept it but want to consider truth relative to correlations that violate it, or by those who want to keep track of where notions like synonymy are being used and where they aren't.

I should note that even those like Quine who reject a notion of interpersonal synonymy grant that there are standards of translation: it is beyond controversy that some translations of a foreign sentence are *better than* others, at least for certain purposes. If someone hears an utterance of "Comment ca va?" and translates it as "You have just passed your French exam", then that grossly violates any reasonable standard of good translation. The skeptic about interpersonal synonymy holds that though we of course have standards of translation, they are a matter of being better or worse, not of right and wrong, and also are highly context-dependent (since the purposes for which they are better or worse vary from one context to the next); and that these features together keep them from generating any useful notion of synonymy. I don't want to discuss whether this rejection of interpersonal synonymy is a reasonable position. But note that even if it is, standards of better and worse translation can be employed to partly capture the import of claims about extended disquotational truth: what we are doing when we conjecture whether some utterance we don't understand is true is conjecturing whether a good translation of the utterance will map it into a disquotationally true sentence we do understand. If "good translation" is a highly context-sensitive and interest-relative notion that doesn't reflect an objective synonymy relation, then there is a certain non-objectivity to the question of whether the utterance we don't understand is true, but that consequence is not obviously undesirable.

(Even a deflationist who thinks there is an objective notion of synonymy for good translations to reflect will have a similar consequence to swallow. For presumably if there is an objective synonymy relation, then there could be utterances that we know don't bear this relation to any utterances we understand, and yet presumably it could still sometimes make a certain amount of sense to discuss whether such an utterance was true. An inflationist can take such a discussion at face value; but as far as I can see, the only way for a deflationist to deal with this, even if he accepts a notion of "objective synonymy", is to appeal to context-sensitive and interest-relative standards of translation that do not reflect objective synonymy.)

In addition to extended disquotational truth and truth relative to a correlation, there is a third option for a deflationist who wants to apply "true" to foreign utterances, and that is simply to use the concept of pure disquotational truth as originally defined in connection with the foreign utterance, without relativization. This may seem odd: didn't I say that my concept of pure disquotational truth applied only to utterances in my own language? No: I said that it was defined only for utterances that I understand. I understand the sentence "Der Schnee ist weiss". Following Quine (1953b, p. 135), I hold that I can apply schema (T) to it, getting

"Der Schnee ist weiss" is true iff der Schnee ist weiss,

which I can perfectly well understand despite its not conforming to the grammatical rules of any standard language. (If it is objected that there would be a problem had I picked a sentence with a different meaning in German than in English, my response is that in that case "Der Schnee ist weiss" would be ambiguous for me, and there would be no problems in this case that don't arise equally for ordinary ambiguity within a language. I will discuss whether ambiguity creates a problem for deflationism in Section 10.)

I don't see any reason to choose between the second and third options: the deflationist can use whichever one suits his purposes of the moment. For some purposes the third option is better than the second. For instance, if I understand a foreign sentence that has no exact equivalent in English, the third option can be used to give more adequate disquotational truth conditions than the second. The third option is also more useful in certain explanatory contexts, when we want to explain a foreigner's behavior without bringing the English language into the explanation. But the second option is often more useful than the third, in that it can in principle be used even for sentences we don't understand.

Sometimes that broader application can come in handy: for instance, in trying to *come* to understand a foreign sentence we may try out several different correlations to our own sentences to see which makes most sense. And in some contexts we want to use the second option in connection with semantic or quasi-semantic notions like synonymy or acceptable translation: for instance, in a perjury trial we might want to know whether there is any acceptable translation of an utterance that makes it come out true. Finally, there are contexts in which the second option is useful even in connection with our own utterances: if we are considering a counterfactual situation in which my use of words is very different, I may be interested in the truth conditions of my sentences relative to a mapping of my language into itself that is not the identity mapping. So a deflationist should not dispense with the second option. In many contexts though I think that the third option is the most natural. It is the one that accords with my original gloss on disquotational truth: a sentence is disquotationally true iff it is true-as-I-understand-it.

I think then that the notion of extended disquotational truth should be avoided if possible: it is not obvious that a deflationist can make sense of it, it is less obvious that he needs it, and it seems better to keep the logical aspects of truth cleanly separated from any semantic and quasi-semantic elements.

9. The modal properties of "true"

A further point about extended disquotational truth as I have defined it is that it is a concept of little theoretical utility. That is because it is a curiously hybrid notion: whether a foreign utterance is true in this sense depends on semantic facts, but whether one of our own sentences is true in this sense doesn't (since as applied to our own sentences, extended disquotational truth is equivalent to pure disquotational truth, which is use-independent). Recall that certain logical uses of the pure disquotational truth predicate "true_{pd}" (for instance, the modal version of the Euclidean geometry example) require that ~("Snow is white" is true_{pd} iff snow is white). So even if "snow is white" had been used in English in pretty much the way that "Grass is red" is actually used, "snow is white" would still have been true in the purely disquotational sense, and hence in the extended disquotational sense as well. Moreover, if in addition grass and the German language had been as they actually are, then "Der Gras ist rot" would have been synonymous with "Snow is white", and hence would have been true in the extended disquotational sense too. The counterfactual supposition of a drastic change in the use of English sentences leaves their extended disquotational truth conditions the same but alters the extended

disquotational truth conditions of foreign sentences! This, surely, is not a theoretically useful way to make truth conditions use-dependent.

Obviously what we would need to do, if we wanted a more useful notion of truth conditions that has semantic features built into it, is to build in such features for our own sentences as well as for the sentences of others. And of course inflationary theories of truth conditions do precisely this. The question is, can one do this on a theory of truth conditions with any claims to being deflationary?

There is an obvious modification of the extended disquotational approach: I don't know if a view of this sort should be called deflationary or inflationary, so I'll call it *quasi*-deflationary. It requires that we have not only a notion of synonymy, but a prior notion of meaning such that two sentences are synonymous if they have the same meaning; but meaning is to be defined independently of truth conditions. (The crude illustration would be that meanings are just verification conditions; slightly less crudely, perhaps meanings are some sort of equivalence classes of pairs of conceptual roles and indication relations; the point of using equivalence classes being of course to allow that small differences of conceptual role don't make for a difference of meaning.) Then we define quasi-disquotational truth (truth_{qd}) and quasi-disquotational truth conditions by:

$$\sim(\text{S is true}_{\text{qd}} \text{ iff } \langle p \rangle \text{ m}(\text{m is the meaning of S and } @(\text{m is the meaning of "p"}) \text{ and } p)),$$

and analogously,

$$\sim[\text{S has the quasi-disquotational truth conditions that } p \text{ iff } \langle \text{m}(\text{m is the meaning of S and } @(\text{m is the meaning of "p"})) \rangle];$$

where "@" is an "actually operator" which "temporarily undoes the effects of" the modal operator.²⁸ This seems to have the modal properties we desire if we want to mimic an inflationary truth predicate in making the truth of sentences use-dependent.

Indeed, it mimics it well enough so that it is natural to wonder whether it shouldn't count as an inflationary notion of truth and truth conditions. For note that according to it, a sufficient condition for having the truth conditions that snow is white is having a certain "meaning" m: perhaps a certain combination of such things as conceptual role, indication relations, and the like. Why then shouldn't this count as an explication of truth conditions in terms of such things as conceptual role and indication

relations?

It does nonetheless seem to me that there may be something to be said for counting a view like this as somewhat deflationary, if the idea of a meaning is explained in a sufficiently "un-truth-theoretic" way. After all, it is compatible with this view that there be no very systematic connection between "meanings" and truth conditions: for instance, it is compatible with this way of defining truth conditions, albeit not very plausible, that the "meaning" of "Snow is white or grass is green" have few systematic links to the "meanings" of "Snow is white", "Grass is green", and "or". Note further that this view assigns truth conditions only to sentences that are synonymous with English sentences (or anyway, sentences that we understand). This points up the fact that there need be no "natural connection" between the "meanings" it assigns and truth conditions: the connection between the meanings it invokes and truth is supplied entirely by the disquotational schema for sentences we understand. So the view is somewhat deflationary, though just how deflationary it is will depend very much on the details of what it says about "meanings".

One kind of worry about views of this sort concerns whether any acceptable notion of "a meaning" and of "having a meaning" will do the job required. Putting aside purely ontological concerns, the main worry concerns the notion of synonymy (having the same meaning) that would be generated by any such view of meaning and the "having" of meanings. Long ago Quine observed that it is an open question whether interpersonal synonymy and intrapersonal synonymy can be viewed as two aspects of the same general equivalence relation. (Quine 1953a, p. 56; see also Quine 1960, sections 9 and 11.) Obviously we employ different criteria in the two cases—for instance, in the intrapersonal case we use a speaker's willingness to substitute one expression for the other, and this criterion is unavailable in the interpersonal case—and reflection on the criteria we do use in the two cases can lead to skepticism as to how well they cohere. I won't try to elaborate this skepticism here, I simply want to note that such skepticism is possible, and that it would make it impossible to regard intrapersonal synonymy and interpersonal synonymy as both cases of the general relation of "having the same meaning". (In addition, the explanation of synonymy as having the same meaning requires synonymy to be an equivalence relation; this is another assumption about synonymy that has been questioned.)

I'm not sure how seriously to take these worries, but as I said in connection with extended disquotational truth, I think it advantageous to cleanly separate the semantic assumptions one is

employing from the logical aspects of truth: failing to do so can easily lead to obscuring the role to which one is putting one's semantic assumptions. (I think in fact there is a serious danger of this in the case of quasi-disquotational truth conditions: it is easy to fall into trying to give them a kind of explanatory role that in fact is precluded given the peculiar way they are defined. But this and other issues about explanation are beyond the scope of this paper.) In any case, I think a full-fledged deflationist should make no commitment to a notion of interpersonal synonymy, let alone to interpersonally ascribable meanings.

Let us return now to counterfactuals like

- (1) If we had used "Snow is white" in certain very different ways it would have had the truth conditions that grass is red.

Presumably the average person would call such a sentence true. Doesn't this show that the average person is clearly not using the word "true" in its purely disquotational sense? And doesn't that in turn show that a version of deflationism that puts purely disquotational truth at the center of things, and which refrains from endorsing intersubjective meanings and hence quasi-disquotational truth, is gratuitously departing from common sense? I don't think so. We could after all say that the "cash value" of (1) is

- (1*) in considering counterfactual circumstances under which we used "Snow is white" in certain very different ways, it is reasonable to translate it in such a way that its disquotational truth conditions relative to the translation are that grass is red;

and that this is all the ordinary person really means to assert. However, I am suspicious of this and all other claims about what the ordinary person means in making assertions involving "true": I rather doubt that there is a consistent way to make sense of all ordinary uses of this notion. I am rather inclined to think that *many* ordinary uses of "true" do fit the purely disquotational mold, though I regard the question of whether this is so as of only sociological interest. If, as may well be true, there is an ordinary meaning of (1) that (1*) fails to capture, then that would show that ordinary speakers are committed to a notion of truth that goes beyond the purely disquotational. But if we can lessen those commitments in a way that is adequate to all practical and theoretical purposes, and if in doing so we can still capture the "cash value" of ordinary utterances as well as (1*) captures the cash value of (1), then the charge that we are "gratuitously departing from common sense" is quite unfounded.

10. Ambiguity and indexicals

So far I have been a bit cavalier about whether a disquotational notion of truth applies to sentence-types or to utterances. I want it to apply to utterances, so that if an ambiguous sentence or a sentence containing an indexical or demonstrative is uttered on different occasions, we can regard some utterances of it as true and others as false.²⁹ One substantial worry about deflationism is whether it can accommodate this.

In dealing with this worry it is important to bear clearly in mind that talk of truth conditions as I am primarily construing it is thoroughly non-semantic. "True" in the purely disquotational sense means "true as I understand it": it doesn't mean "true on the correct understanding of it", because the idea of a "correct understanding" of a sentence or utterance is a semantic notion that has no place when we are discussing purely disquotational truth. If on my understanding of "Der Schnee ist weiss" it is equivalent to " $E=mc^2$ ", then for me this sentence is disquotationally true iff $E=mc^2$.³⁰ Anyone in the grip of the Frege-Russell tradition will think that this shows that we need a notion of truth conditions very different from the disquotational one; but I don't think it at all obvious that they are right. As I observed in Section 8, a deflationist can grant that there are standards of adequacy of translation that do not presuppose truth conditions, and can grant the legitimacy of saying that relative to a translation manual that meets those standards the sentence will have different truth conditions. So a notion like *A*true as I ought to understand it (given my goals of translation together with the facts about how the sentence is used)@ certainly make sense; it just needs to be distinguished from the purely disquotational notion that I am taking to be primary.

Can the purely disquotational notion be applied to ambiguous utterances and indexical utterances? Let's start with ambiguity. I believe that it poses little problem for the deflationist. There is a rather common-sensical story of how we process ambiguous sentences, which I assume the deflationist can help himself to: the story has it that when we come upon a two-way ambiguous sentence like "Visiting relatives can be boring" or "I met my lover at the bank", there are two alternative ways of processing it and using it in inferences, one corresponding to each of the interpretations; indeed, interpreting it one way rather than the other just is processing it in one way rather than the other (for instance, storing it in memory in such a way that it will be used in inferences of one kind rather than the

other). When we are undecided as to how to interpret an ambiguous sentence, what we are undecided about is how to process and store it and which of the two sorts of inference to make with it. The fact that we can process and store a sentence in two different ways can be put metaphorically by saying that we attach "inner subscripts" to lexically ambiguous elements and "inner syntactic markers" to mark different ways of resolving the syntactic ambiguities. This metaphor allows us to talk, if we like, as if each person thinks in an ambiguity-free language. On this way of talking, to say that an utterance is disquotationally true is to say in effect that the sentence in my ambiguity-free inner language which I associate with the utterance is disquotationally true. Of course, if I am undecided how to interpret the utterance, there is no unique ambiguity-free inner sentence to use and hence no unique disquotational truth conditions will be generated. We can regard that as a case where we don't understand the utterance, so that the notion of disquotational truth doesn't apply. Although of course we can also ascribe it truth conditions relative to each of the ways of understanding it. The metaphor of the ambiguity-free language is not really playing an essential role here: I could equally well have said that I accept all instances of "That utterance of 'Visiting relatives can be boring' is true iff visiting relatives can be boring", but may process the right hand side in either of two ways on different occasions. (I disallow inference from the right hand side processed one way to the right hand side processed the other way.)

Now let's turn to sentences containing indexicals and demonstratives. Here it is natural to divide up the account of truth into two stages. The first stage concerns sentence types: here of course there is no hope of defining an unrelativized truth predicate, since sentence-types like "I don't like her" don't have a truth value, but we can associate a truth value to this type relative to a pair of objects $\langle b, c \rangle$; to say that "I don't like her" is true relative to $\langle b, c \rangle$ is simply to say that b doesn't like c . (I suppose that this isn't strictly "disquotational" in that it involves a grammatical adjustment from "don't" to "doesn't", but surely we can allow the disquotationalist the machinery required for this grammatical adjustment.)

The second stage is to provide an account of unrelativized truth for sentence-tokens. Given the first stage, this simply amounts to associating objects with each of the indexical elements in the sentence, for then we can declare the token true iff it is true relative to the sequence of associated objects. How are we to do the association? If we focus only on indexicals like "I" and "now", there is a simple rule that would work pretty well: always relativize "I" to the producer of the token and "now" to the time of production. But this rule isn't invariably such a good one: consider the subway advertisement "Stop smoking now". At any rate, there seems to be no such simple rule for "she", or "that": the "correct"

object to assign, if it makes sense to talk of correctness here, depends in very complex ways on the intentions of the producer. In these cases anyway, and probably even in the cases of "I" and "now", we must regard all talk of the "correct assignment" as a semantic matter which cannot appear in a purely disquotational account of truth conditions.

Still, nothing stops us from applying talk of disquotational truth conditions to tokens: once we remember that "disquotationally true" means "true-as-I-interpret-it", the obvious thing to say is that an utterance of " $p(i_1, \dots, i_n)$ " is disquotationally true (for me, that is, as I understand it) iff the sentence is true relative to the values of a_1, \dots, a_n I regard as appropriate to associate with the indexicals.³¹ When I say that I "associate values" with an indexical, of course, what I do is associate a mental occurrence of one of my own expressions (possibly itself indexical) with it. If I can't associate a term with an indexical in a sentence, then I can't attach disquotational truth conditions to the sentence. This is just an extension of a point made earlier for non-indexical utterances: sentences that I don't understand have no disquotational truth conditions for me.

I've said that the expression I associate with an indexical may itself be indexical: I may assign disquotational truth conditions to an utterance of "She is here now" by thinking "She is here now". As with ambiguity, the deflationist needs to say something about how thoughts involving indexicals are processed, but the story is very similar to the story for ambiguity. When I think a thought involving "she" to myself on a given occasion, that thought will typically hook up causally to a certain "internal file drawer" of thoughts involving other singular terms, perhaps "Sheila" and/or "that woman I saw at the beach last Friday". (It is possible that all the singular terms in the "file drawer" are indexical: you can use an indexical to "open a new file drawer". Even when there are other terms in the "file drawer", nothing need pick out any one person uniquely: the function of these associated terms isn't to explain how we can be referring to a certain person, if indeed we are; the function is simply to tell an obvious story of what it is for me to *regard* a given occurrence of "she" as "about Sheila" or "about that woman I saw at the beach".) Again, you could if you like say that the internal occurrences of indexicals are "subscripted" so as to remove their indexicality; I take this to be just a dispensable manner of speaking for something better stated in terms of the internal processing.

We see that there is no special difficulty in explaining disquotational truth conditions for indexical tokens. If indexicals raise any special difficulty for deflationism, it is that for indexicals it is less believable

that we don't need a more inflationist notion of truth conditions. Surely, it may be said, there will typically be a *correct answer* to the question of who another person was referring to with a particular application of "she"; and the deflationist seems unable to accommodate this.

It seems to me though that the internal processing story does a lot to accommodate it. In a typical case where we misinterpret a token of "she" where we incorrectly interpret the speaker as meaning Mary when in fact she meant Sheila we do so because of false beliefs about the speaker's internal processing: we think that her token was connected up to an internal file drawer of thoughts involving terms like "Mary" when in fact it was connected up to a file drawer involving terms like "Sheila". Standards of appropriate translation are going to rule that in that case, "Mary" is a bad translation of "she". The sort of facts that this brings into the standards of good translation are not in themselves semantic, so it seems hard to argue that the deflationist isn't entitled to appeal to them. Of course, it may be possible to argue that when we describe the standards of acceptable translation for indexicals in detail we will have to bring in machinery that is powerful enough to provide a reduction of the semantic notion of reference to non-semantic terms; if this is so, then the would-be deflationist is in fact turning himself into a reductionist inflationist. As I remarked earlier, the distinction between deflationism and reductionist inflationism may not in the end be altogether clear. At any rate, it certainly isn't obvious without argument that the standards of acceptable translation for indexicals give the machinery necessary to reduce the semantic notion of reference to non-semantic terms.

11. Learning from others

The final objection that I will consider is that you need an inflationist view of truth conditions to make sense of how we learn from others. Much of our information about the world is acquired from others in particular, from their utterances, and from the beliefs we infer them to have. The notion of truth figures in this process: typically, what we do is assume that certain things the other person tells us are probably true or that certain things we can see that she believes are probably true. It might be thought that an understanding of this process requires a non-deflationary notion of truth; or at least, that it requires more than purely disquotational truth. There are actually two distinct lines of argument that might be given for this conclusion.

The first is that since we learn from other people, including people who speak different

languages, translation must somehow be involved, so that some notion of interpersonal synonymy is presupposed in the inference. If this were right it would show that pure disquotational truth, without a synonymy relation, was inadequate for these purposes: you'd need at least extended disquotational or quasi-disquotational truth, and maybe a more thoroughgoing inflationary notion, for the purpose. I'll get back to this after discussing the other line of argument.

The second line of argument purports to show that even pure disquotational truth plus synonymy isn't enough. Let's slur over the possible need for a synonymy relation, by imagining that I am learning from Charley, a speaker of my idiolect. Charley says: "You know, on one day in 1936 there was over a foot of snow on the ground in Mobile Alabama". Do I infer that one day in 1936 there was over a foot of snow in Alabama? For most Charleys and most circumstances, surely not. But suppose that in the past Charley has asserted each of the following: "You know, there are parts of Virginia that are north of parts of New Jersey"; "You know, in 1928 the Soviet government secretly supplied arms to Chiang Kai-Shek to be used against Mao"; "You know, Leon Russell was one of the studio musicians for the Fleetwoods". Each time I've found the claim almost unbelievable, but on checking it has turned out true. If he has never said "You know, p" for a p that has turned out false, then I think I might well believe his claim about Mobile.

A crude way to formalize my inference is as an enumerative induction: all the sentences of a certain sort that Charley has uttered under certain circumstances that have been independently checked have been true; "There was a foot of snow in Mobile" is a sentence of that sort uttered under those circumstances; so (in the absence of defeating evidence) probably that sentence is true. In addition to the inductive inference, we need an instance of the disquotation schema, to get from the claim about truth to the object level claim. Obviously the disquotationalist has no trouble with the last step. But the inductive part of the inference might seem more questionable. For suppose we were to try to formalize the inductive inference without use of the word "true". What we would get is this:

1. There are parts of Virginia that are north of parts of New Jersey.

....

n. In 1928 the Soviet government secretly supplied arms to Chiang Kai-Shek to be used against Mao.

n+1. Charley said "You know, there are parts of Virginia that are north of parts of New Jersey."

....

2n. Charley said "You know, in 1928 the Soviet government secretly supplied arms to Chiang Kai-Shek to be used against Mao."

2n+1 . Charley said "You know, on one day in 1936 there was over a foot of snow in Mobile Alabama."

Therefore (probably): On one day in 1936 there was over a foot of snow in Mobile Alabama.

And a possible reaction to this inference so formulated is that it has no inductive force whatever: none of the premises, have anything to do with snow or Mobile; the first n have to do with diverse topics, and the last n+1 have to do only with noises that Charley uttered. So, it might be argued, when we reformulate the premises and conclusion of the inference in terms of truth, this can't be the innocent reformulation that the disquotationalist says it is: for the displayed inference has no inductive force, but the reformulation in terms of truth does have inductive force. So the reformulated inference works by attributing a more substantive property than disquotational truth to Charley's utterances, a substantive property on which it is legitimate to enumeratively induce.

I have tried to set out this line of argument persuasively, but in my view it rests on a very naive theory of induction. (I'm not talking about its use of enumerative induction: I take the assumption that learning from Charley is based on an enumerative induction to be a simplifying assumption on which nothing important is likely to turn.) I grant that to someone who had no concept of truth whatever, the displayed inference would probably have no inductive force. But that doesn't imply that we who do have a concept of disquotational truth should attach no inductive force to the very same inference (or equivalently, to the inference as reformulated in terms of disquotational truth). To be sure, our attaching force to this inference does not inevitably follow from our having a disquotational truth predicate; in principle, someone could perfectly well have a disquotational truth predicate and regard it as "unprojectable". But the fact that we aren't like that, that we do regard it as "projectable", doesn't seem to me to have much to do with whether the predicate is "substantive" in the sense employed in the objection: it doesn't show that "true" isn't just a logical predicate defined by its role as a device of infinite conjunction. (It's worth noting that someone could equally well introduce a more theoretically loaded

truth predicate and find it "unprojectable": "grue" after all is a "substantive" predicate in the sense in question here, that is, it is not a mere logical device in the way that disquotational truth is.) Our acceptance of enumerative inductions on "true" is presumably due to having discovered that we get useful results by doing so,³² and such a discovery does not require that truth be anything more than what the disquotation schemas tell us.

If this is right then the second argument that learning from others is incompatible with deflationism has no force. How about the first argument, that learning from others requires an interpersonal notion of synonymy, so that pure disquotational truth by itself is inadequate to the job? This too seems wrong. Suppose Charley speaks another language, which I think I understand: anyway, I have a translation manual that I use for it. Suppose I have found out in the past that when Charley utters sentences, they usually turn out to be true as I understand them—that is, their translations turn out true, on the translation manual I use. Then we have an inductive argument that the next one will be true, on the manual I use: in this way, we can inductively argue to the snow in Mobile, as before. As far as I can see it doesn't matter in the least whether this manual accords with the genuine interpersonal synonymy relation between his sentences and ours, assuming that such talk of the correct interpersonal synonymy relation even makes sense: truth-relative-to-the-manual is the only notion we need in learning from others.

12. Conclusion

I repeat that there are many other arguments to be considered for the need of a more-than-disquotational notion of truth or truth conditions, some much more complicated to deal with than any given here. What I have tried to do here is simply to motivate a fairly strong version of deflationism—strong in that it does not rely even on a semantic notion of synonymy—and to sketch how such a version of deflationism can overcome various objections. I think what I have done here makes some case for what I've called "methodological deflationism", the idea that we should assume full-fledged deflationism as a working hypothesis. That way, if full-fledged deflationism turns out to be inadequate, we will at least have a clearer sense than we have now of just where it is that inflationist assumptions about truth conditions are needed.³³

Footnotes

1. Because of the paradoxes, exceptions must be made for certain utterances u that contain "true"; I won't be concerned here with just how the exceptions are to be carved out.

2. There is more than one acceptable way to understand "cognitively equivalent" here. My own preferred reading, for what it's worth, is that to call two sentences that a person understands "cognitively equivalent" for that person is to say that the person's inferential procedures license a fairly direct inference from any sentence containing an occurrence of one to the corresponding sentence with an occurrence of the other substituted for it; with the stipulation, of course, that the occurrence to be substituted for is not within the context of quotation marks or an intentional attitude construction.

(Cognitive equivalence *relative to some other assumption* is the same, except that that assumption is allowed to be used in the inferences.) I would also take the claim of cognitive equivalence to imply that the inferences are more or less indefeasible. (More specifically, that they are empirically indefeasible, and close to indefeasible on conceptual grounds as well, and that the person is not in possession of defeaters for them. These stipulations are motivated by the assumption that we should describe someone who doesn't know the semantic paradoxes as a person for whom "'p' is true" is cognitively equivalent to "p" across the board; as the person comes to terms with the paradoxes he revises his standards of cognitive equivalence on conceptual grounds.)

3. Certain uses of truth conditional talk in explanations seems to me the most likely target for excision.

4. More radical, I think, than the one advanced in Horwich 1990. While I like much that Horwich says I have substantial disagreements: see Field 1992b.

5. Of course, conceptual or computational role can be described at different levels of abstraction, so

there is more than one notion of conceptual role that meets this description. I think the notions of meaning and content are too indeterminate for there to be a right answer about which notion of conceptual role to include; one could even include more than one.

6. If the probability of S given E is higher than the probability of S, then (especially if this fact is rather "robust", that is, if except for very special sentences F, $P(S|E\&F)$ remains higher than $P(S|F)$) it is natural to include this fact in the conceptual role of S, even if E is in no sense "observational" and could never be established with enough certainty or whatever to count as evidence for anything. It was central to verificationism to exclude such facts from verification conditions when E was not observational.

7. Part of the idea behind this assumption is that when you read or hear a given unambiguous sentence, there is a type of thought state that you typically undergo in processing it, where this type is identifiable in purely computational terms; and that this type of thought state also typically occurs when you utter such a sentence. This coupled with some other plausible assumptions about the computational processes of the language-user make it possible to define in computational terms a correspondence between many of the person's states of believing and desiring and their unambiguous sentences. A natural computational story about the processing of ambiguous sentences allows us to extend this to such sentences too. I will not attempt to give the details of any of this here.

8. The deflationist can also formulate the distinction for *other peoples'* belief states, by relying on translation, though to the extent that that is indeterminate the distinction there can obviously be given no objective significance.

9. I have heard the objection that a deflationist is free to give truth conditions an explanatory role as long as it is only "deflationary truth conditions" and not "heavy duty truth conditions" that are given that role. This seems to me a confusion. I think it is quite misleading to speak of two kinds of truth conditions, the

heavy duty ones and the disquotational ones: this suggests that the advocate of "heavy duty" truth conditions thinks that we should ascribe to "Snow is white" some truth conditions other than that snow is white, and that of course is absurd. It would be better to say that there are two kinds of relation between truth conditions on the one hand and utterances and mental states on the other, the disquotational relation and the heavy duty relation. But the only ways I can see to make sense of the distinction between these relations directly or indirectly preclude giving the disquotational relation certain kinds of explanatory role. (In my view, the preclusion is indirect, and stems from the deflationist's account of how truth conditions attach to sentences and mental states.)

10. This last way of putting things gives a way to make content social without relying on an independent notion of meaning. It thus leaves room for explaining meaning in terms of content if so desired.

11. I mention "not" because the deductive inference rules I listed include the rule that we can infer "not (A or B)" from "not A" and "not B" together. It is possible to replace this by a rule involving "if...then" instead of "not"; or (as in natural deduction systems and sequent calculi) by a rule involving a notion of implication.

12. It may not avoid the threat of indeterminacy in the end: I discuss this in Chapter 8. Even so, the account to follow offers an attractive alternative to the inflationist's approach to the determination of the way that our logical symbols contribute to truth conditions; and I will generalize it in Section 6.

13. The argument from (T) to (TF) generalizes as an argument from (TG) to

$$(p \text{ (} q[\text{"p or q" is true if and only if "p" is true or "q" is true}].$$

By use of the additional axiom

$\forall x[x \text{ is a sentence of our language} \supset p(x="p")]$,

we get (TFG).

14. Or at the very least, description theories that use only non-metalinguistic descriptions.

15. This incidentally is why the Benacerraf problem about mathematical knowledge (Benacerraf 1973) isn't dissolved by a disquotational theory of truth. [See the reformulation of the problem in Section 5 of Chapter 11.]

16. Methodological deflationism doesn't even preclude that we might in the end come to accept an *unreduced* inflationist relation *S has the truth conditions that p*: we might conceivably see the need for introducing non-physical relations as explanatory. I think it does put the burden of proof against such a position, but I think that doing so is appropriate.

17. This needs two small qualifications. The first is that the generalizations can be formulated otherwise if the language contains certain devices of infinite conjunction such as substitutional quantifiers. However, this qualification is not a terribly severe one, since the presence of such substitutional quantifiers would allow you to define "true": S is true iff $(\exists p(S="p" \supset p))$. (Again, I assume a theory of substitutional quantification that avoids the paradoxes.)

Note that the use of schematic letters as an alternative to substitutional quantification, mentioned in Section 3, does not allow us to formulate infinite disjunctions without "true", and therefore does not allow us to explicitly define "true".

The second qualification is that the view I've given (in either the substitutional quantifier version or the free schematic variable version) allows statements that are really a bit stronger than infinite conjunctions, in the same way that first order quantifications are stronger than the totality of their

instances even when every object has a name: see the discussion in Section 3 of how to get the generalization (TFG).

18. For present purposes I am regarding sentence types as "necessary existents", so that the fact that the disjuncts don't all entail the existence of sentence types doesn't debar a sentence that does entail this from counting as their "infinite disjunction".

19. Both Hilary Putnam (1983, pp. xiv-xv and 279-80) and Crispin Wright (1992, pp. 12-24) object to deflationism on this ground. However, their discussions are directed at a version of deflationism which says that truth is not a property (Putnam) or not a "substantial property" (Wright), and I'm not clear enough as to what that is supposed to mean to know whether the authors would have intended to be arguing against deflationism as I have defined it. (In a long footnote, Wright does say that he thinks his objection applies to the version of deflationism in Horwich 1990, despite the fact that Horwich disavows the "no property" claim.)

20. Opponents of deflationism sometimes try to pin on it the claim that when we come to understand new sentences, or when new words and hence new sentences are added to our language, our concept of truth changes. But a deflationist can agree that on the most natural ways to individuate concepts over time, adding new sentences to the language (or to our domain of understanding) leaves the concept of truth unchanged: after all, once we have come to understand the new sentences, the truth schema dictates how the word "true" is to be applied to them. This doesn't conflict with what is said in the text: the point there was that we don't now understand attributions of disquotational truth to sentences containing future words. (Also, there is no reasonable sense in which such attributions now have disquotational truth conditions, since what truth conditions they may come to have will depend on how the future words will be used.)

21. We can introduce a purely disquotational notion of what it is to express a proposition, with the property that claims of the form "'p' expresses the proposition that p" are necessary truths. We can also introduce certain impurely disquotational notions of what it is to express a proposition; these are analogous to the impure disquotational views of truth that we will consider later. Deflationists about meaning-that and believing-that think that such disquotational notions of expressing are the only legitimate ones; inflationists disagree. This distinction between inflationist and deflationist notions of expressing a proposition obviously needs to be clarified, if one is going to talk in terms of propositions; I think that some of the rest of the paper suggests a way to clarify it, but I will obviate the need to discuss this by avoiding talk of propositions.

22. Or a more basic notion like satisfaction from which truth is defined.

23. Actually even the demand of a compositional *inductive definition* must be relaxed, if we want to both get a notion of truth that applies to sufficiently many sentences that contain semantic terms and simultaneously avoid limiting the truth predicate to sentences containing no unrestricted quantifiers: the usual methods like Kripke's for defining a notion of truth that applies to sufficiently many sentences that contain "true" require a restriction of quantifiers *even in the inductive definition* (not just in turning it into an explicit definition); that is, what is being defined isn't truth but only "comes out true under such and such a restriction of the quantifiers". I think that a lesson of the paradoxes is that we must go to an axiomatized truth theory rather than defining "true", even inductively.

But this doesn't undermine the point intended in the text: for one might well demand (more or less in the spirit of Tarski or at least Davidson) that an axiomatized truth theory be built around explicit compositional axioms like "A disjunction is true if and only if at least one disjunct is true". A more full-fledged deflationist view would be that such compositionality has no intrinsic importance.

24. Let P be a one place predicate and N a name. By the assumptions indicated, ' $F(P="F")$ ' and ' $a(N="a")$ '. By laws of concatenation and quotation marks, it follows that ' $F' a(P^{\wedge}N="Fa")$ '.

By syntax, $(F(a["Fa"] \text{ is a sentence}))$, so by the truth schema, $(F(a["Fa"] \text{ is true iff } Fa))$. By first order logic, we can rewrite this as $(F(a["Fa"] \text{ is true iff } \exists x(Fx \text{ and } x=a))$. But then the appropriate schemas for truth-of and reference will allow us to rewrite this as $(F(a["Fa"] \text{ is true iff } \exists x("F" \text{ is true of } x \text{ and } "a" \text{ refers to } x))$. Combining with the previous paragraph, we get that $P^{\wedge}N$ is true iff $\exists x(P \text{ is true of } x \text{ and } N \text{ refers to } x)$.

The step labelled "by first order logic" implicitly assumes that N denotes, that is, that ' $a[N="a"]$ ' and $\exists x(x=a)$. If this supposition is abandoned, some form of free logic must be used; the details of the free logic affect the details of the compositional principle that is derived.

25. At least, this is so for what I've called a "pure deflationist". The idea of "impure deflationism" will be considered in Section 9.

26. I made a rather confusing attempt to state the problem in Field 1986, Sections IV and V. I now think there is a way around the problem there raised. (Stephen Leeds helped set me straight on this: see Leeds 1995, a draft of which I received just as this paper was being completed.) [For further discussion, see Section 7 of the Postscript.]

27. I'm inclined to think that the problem isn't just that interpersonal synonymy ("is a good translation of") is a highly vague notion, but that it isn't a fully factual notion at all; rather, it is an evaluative notion.

This is compatible with many claims of interpersonal synonymy being factual relative to the goals of the translator; but not all interpersonal synonymy claims need be factual even relative to the goals. (For more on the idea of evaluative claims as *not fully factual*, see Chapters 8 and 13.)

28. See for instance Crossley and Humberstone 1977.

Quantification over meanings seems necessary in this definition: certainly you can't replace the right hand side by " $\neg p \leftrightarrow (S \text{ is synonymous with } 'p')$ and p ", since this makes only the actual meaning of S relevant to whether S would be true in a non-actual context, contrary to intentions. And you can't simply take meanings to be equivalence classes of expressions under the (actual) synonymy relation: for then the proposed definition would collapse into the inadequate one just mentioned. In these remarks I'm assuming that you don't quantify over possible worlds: if you do, then you can use the relation *expression e as used in world w is synonymous with e' as used in w'* , and take meanings to be equivalence classes of pairs of expressions and possible worlds under this equivalence relation.

29. Incidentally, I remind the reader that I am counting it as an example of ambiguity if someone understands the same sentence as meaning one thing in German and something else in English: this was required by my adopting the "third option" for dealing with sentences in other languages, discussed in Section 8.

30. Since it is disquotationally true iff *der Schnee ist weiss*, and hence, by the equivalence hypothesis, iff $E=mc^2$.

31. Typically these values will be the ones I take the producer of the utterance to have intended, but not always: e.g., "now" on the sign intended to be read by various readers at different times. And since we're talking disquotationally true for a hearer, the actual intentions of the producer don't matter, what

matters is the reading the hearer gives.

32. We may well also have an innate predisposition to believe other people, barring evidence to the contrary; this does not affect my point.

33. I've had helpful conversations with many people about these issues in recent years. Among those whose comments have influenced my presentation are Marian David, Michael Devitt, Paul Horwich, Barry Loewer, and Stephen Schiffer. I have also incorporated some useful suggestions by the editors.