ACTA UNIVERSITATIS STOCKHOLMIENSIS

Stockholm Studies in Philosophy 29

DEFLATIONISM:
A USE-THEORETIC ANALYSIS
OF THE TRUTH-PREDICATE

ARVID BÅVE

Stockholm University

TABLE OF CONTENTS

Preface

When asked about the topic of my doctoral dissertation, the little word "truth" is normally a guaranteed conversation killer. For some reason, it has come to be surrounded with such an air of grandiose solemnity that people appear to *hear* it spelt with capital "T". The theory I am about to defend of course gets its name from its aim of *deflating* this word, showing how simple and practical its central workings and purposes. It is about truth with lower-case "t". In the end, however, the details of this linguistic phenomenon bring into view something that I find ever more fascinating than the traditionally envisioned grand connection between us and the world. But the fascination is like the one evoked by a simple and ingenious solution to a practical problem, like the wheel or computer mouse, or, perhaps better, the camouflage of the wandering stick. This fascination with linguistic matters is perhaps the irrational reason for my conviction that many philosophical issues essentially turn on linguistic ones, in particular, issues surrounding truth. More rational reasons, hopefully, are the arguments that comprise most of this book.

Many graduate students experience a notorious oscillation between megalomania and self-critique, but in the present case, I fear there may have been too little of the latter. For correcting this imbalance, I owe much to my supervisor Peter Pagin, without whose wide competence and articulate argumentation this dissertation would have been far less readable and persuasive. I am also grateful to Paul Horwich, who kindly invited me to the City University of New York as visiting scholar in Fall 2004, and for valuable comments on Chapter 3. Likewise, I thank Ian Rumfitt for reading and commenting on an early draft of Chapter 4, and for agreeing to be my academic advisor during my stay at Oxford University in Hilary Term 2004. Thanks are due also to Tor Sandqvist of Umeå University, who was responsible for the pre-defence seminar. I am also indebted to the participants

CHAPTER ONE:
AN INTRODUCTION TO DEFLATIONISM

1.1 WHAT IS DEFLATIONISM?

If one were to respect everyone's use of the term, the only answer to this question would have to be, "Nothing very determinate". It is a theory about truth that aspires to explain what truth is, or to characterize the concept of truth. From here, however, terminologies begin to diverge. But we must start somewhere, and the majority, at least, would agree that deflationism takes the following kind of claim to be more or less sufficient for explaining truth: *the proposition that snow is white is true if and only if snow is white*, and analogously for every other proposition. For those who take truth to be a property of sentences, the biconditional is: *the sentence "Snow is white" is true if and only if snow is white*. It should be emphasized that the deflationist does not merely accept these biconditionals – they are accepted by almost everyone – but believes in addition that they give an exhaustive account of truth, in the sense that they suffice for explaining everything a truth-theory should explain. Because the biconditionals are in themselves rather uncontroversial, the dialectics surrounding deflationism is therefore mainly focussed on what the biconditionals can and cannot do by way of explanation, rather than on finding alleged counter-examples to the theory. Accordingly, deflationists do not attempt to find linear arguments for their theory, but argue, rather, by inference to the best explanation.

Deflationism is typically coupled with a claim about the *raison d'être* of the concept of truth, being thus simple to characterize. One might wonder why we have such a concept, and, more importantly, why it has attracted such attention in philosophy, and made such a stir in general, if these trivial biconditionals are all there is to it. The idea is that a word whose meaning is

given by such equivalences enables us to express certain types of generalisa-
tions that we could not express without it. For instance, the command "Tell
only the truth!" is, by the deflationist's light, no more than the command to
say that snow is white only if snow is white, to say that grass is green only if
grass is green, and so on *ad infinitum*. The sentence "Everything he said is
true", furthermore, entails every sentence of the form "If he says that p, then
p", e.g., "If he said that snow is white, then snow is white". Uttering this
universal sentence is therefore a way of saying everything that is expressed
by sentences of the form "If he says that p, then p". Because the instances of
generalisations involving "true" are sentences which vary in certain sentence-
positions, one may say that "true" enables us to *quantify into sentence
position*. The importance and fundamental character of truth, not to mention
the solemnity surrounding it, is thus meant to derive from the fact that there
are many important things that can only be said if we have a word with this
type of generalising property.

What mainly motivates deflationary theories of truth is, first, the idea that
no more than the above equivalence, or something like it, needs to be
assumed about truth in order to explain the related phenomena, and that,
because of its simplicity, the theory is therefore more plausible than any
alternatives. The theory also has the advantage of avoiding various prob-
lematic explanatory commitments of other truth-theories, in particular, to
explain in a substantial way what facts, or states of affairs, are, and to explain
various representational, or "correspondence", relations between truth-
bearers and facts or other objects in the world.

Theories of this spirit have also been labelled *Redundancy Theory*,
Minimalism, *Disquotationalism*, *No-Truth theory*, *Disappearance theory*,
"Ditto" theory, and more, but these labels have not been used consistently to
pick out individuating properties of theories in a systematic way. Many, but
mostly critics, have held that deflationism essentially claims that there is no
property of truth. This is denied by most deflationists themselves, however,
and should therefore not be taken as defining deflationism. It is clear that one
can only speak of a family of ideas here, because of the extent to which the
various accounts differ, both in wording and regarding the interests and
emphases of their originators. I am afraid much ink has been wasted due to
the terminological confusion surrounding these labels. Another phenomenon
detrimental to the debate has been the tendency of describing deflationism in
metaphors and even outright falsities. According to deflationism, it has been
said, truth is "metaphysically thin", "flat", "insubstantial" and even "uninter-

esting". Of course, these phrases serve badly as clarifications. Though I have felt the attraction of some of them as introductions to, or summaries of, deflationism, the dialectical effects of using them make it obvious that they should never have been associated with the label in the first place. To get clearer about this family of theories, we will in 1.4 list a number of theses commonly associated with "deflationism".

One thing that can safely be assumed about all deflationists, however, is that, being ists, they believe their ism is *the* right one. But one may differ on what it is for one true claim about something to be the correct theory about it, rather than merely true. The critical examinations of other truth-theories in the following two sections will contain arguments to the effect that though the claims of their theory may be *true*, they do not constitute *the* correct theory, so I should first say something about this notion of *the* correct theory about something.

Intuitively, one can say that God believes all and only the true propositions, without thereby intending to say what truth is, but only to describe God. So we can at least make some pretheoretic sense of the distinction in question. It also seems plausible that there is an explanatory hierarchy among the true claims, and so, among the true claims about truth. Otherwise, every true claim would be equally primitive. Whether I am right or not that deflationary equivalences are to be taken as primitive is debatable, but not, I believe, that theories of truth should be competing for this status of primitiveness.

There is now a stronger and a weaker conception of "correct theory" to distinguish. First, one can use this notion as standing merely for the "best" theory without taking truth-theories to have any other purpose or standard than satisfying general desiderata like simplicity and explanatory scope. Second, one may hold that one claim about truth is the correct theory in a more objective sense, i.e., that there is something over and above its being "best" which distinguishes it from other claims and makes it the correct theory. The proponent for the stronger view will plausibly take the satisfaction of desiderata as evidence that the theory is correct in the objective sense. Since deflationists take (a generalisation of) the truth-equivalences to be explanatorily exhaustive (i.e., all other truth-facts can be explained by it), they will presumably take it to be the best theory, since there is hardly a simpler one.

The reason that I take the objectivist stance, and why I believe that the best theory is that which says "what truth is", or "explains the concept of

truth" will emerge in Chapter 3, where the correct theory of truth is taken to be in fact an empirical theory about the semantics truth-predicates in natural languages. But this identification of "the best theory" with "the theory which says what such and such is", is rather widespread, if seldom stated explicitly. After all, our only reason for believing that heat *is* mean kinetic energy is that the theory that says it is, is the best one. This inference may even be constitutive of our theoretic faculty, perhaps like our disposition to believe the simplest theory. One can also have a non-realist theory of what one is characterizing, while still distinguishing the correct theory of the concept from merely true claims involving it. For instance, a non-realist theory of evil would not take the claim that G. W. Bush thinks abortion is evil to explain the notion, even if it is true. Therefore, an objectivist view (in the sense above) of this matter does not entail a realist view of the notion to be explained; it may equally reflect the opinion that the theory is to register an objective fact of language use.

1.2 DEFLATIONISM VS. ALTERNATIVES, PART 1: EPISTEMIC AND PRAGMATIC THEORIES OF TRUTH

This section and the next are intended to situate deflationism among its rivals. A rough taxonomy of truth-theories would most plausibly put deflationism in the same camp as correspondence theories, because they are both, well, non-epistemic and non-pragmatic. Very broadly speaking, they both take truth to be a kind of agreement with reality, with no human intervention. Because of their basic agreement, correspondence theorists and deflationists have pretty much the same main objections to epistemic and pragmatic theories. These will be presented in broad strokes in this section, while the next section compares deflationism with correspondence theories. These sections are not mainly intended to persuade those familiar with these issues, but, rather, to show what deflationists (and others) typically have found unattractive about the alternatives, and, occasionally, why deflationism has been thought to come off better in the relevant respect. In addition to this, I will present some comments and arguments of my own, which will appear as I have seen fit in connection to the familiar arguments.

The traditional theories that will concern us in this section are, firstly, the *Coherence theory*, which holds, roughly, that something is true iff it coheres

with (most of) the beliefs that are held (or ought to be held); second, the *Pragmatic theory*, which comes in two variants: one says that something is true iff it is what we come to believe after sufficient examination and scientific theorizing (Peirce (1878: 206), Dewey (1938: 345)), the other takes something to be true iff belief in it promotes practical success (James (1907)); thirdly, *Putnam's theory*, which holds that something is true iff it would be verified in an epistemically ideal situation (Putnam (1981: 54ff.)); fourthly, *epistemic theories about mathematical truth*, according to which a mathematical statement is true iff there is a proof for it.

The major drawback of James's Pragmatic theory and the Coherence theory is that there seem to be clear counter-examples to their respective truth-analyses. To wit, it seems fully possible to have a false belief that for some reason promotes practical success. It also seems at least conceptually possible to have a false belief that coheres with (most of) the beliefs one has. This is not possible if it is impossible to have mostly false beliefs, of course, but this is quite controversial. Further, for many propositions, neither they nor their negations cohere with people's set of beliefs, e.g., the proposition that there is an even number of cars in the world. But one of them must be true. It is common to regard these theories as results of conflating the question what it is for something to be true with that of what makes one justified in judging it true (for an elaborate discussion, see Kirkham (1992: Ch. 2)). Deflationists and correspondence theorists, who take themselves to be competing mainly for best accommodating the data, tend to regard these theories as non-starters in their outright contradiction with our intuitions.

Peirce's, Dewey's and Putnam's theories, though also epistemic in character, are more difficult to assess in that the "ideal epistemic situation", or "sufficient examination and scientific theorizing" have not been spelled out. It has been suspected, however, that once this is done, the theory in question will either be subject to counter-examples or to a charge of triviality (if the notions are defined so as trivially to guarantee the truth of the analysis). Concerning "ideal epistemic situation", for instance, take the simple empirical statement that this wall is white. What is an ideal situation for the verification of this statement? It can hardly be one in which one is looking at the wall, for one may be having a non-veridical perception. This holds also when the lighting is "normal", and, of course, this notion would also have to be defined without trivialising the analysis. It is difficult to see how "epistemically ideal situation" could be spelled out to give enlightening and correct truth-conditions for "This wall is white". It has to be spelled out so that it

comes out true that *if* I am in such a situation and (justifiably) believe that *p*, then *p*. Similarly, it is difficult to see why scientific theorising should necessarily lead us closer and closer to the truth, unless such theorising is simply defined so that this is trivially entailed. But perhaps these ideal limits could be taken as not so abstractly ideal after all. Perhaps they can be explained in terms of intersubjective agreement among rational subjects, or some such. Then, however, the deflationist/correspondence theorist will complain that the theory is only a slight improvement of the former, more radical kind of epistemic/pragmatic theory. It will still deem intuitively possible situations to be impossible, for instance a situation in which all rational subjects come to believe something false. The task for these theories in general can be said to be the finding of a balance between triviality and implausibility.

It has seemed that truth for mathematical statements is more easily analysed in epistemic terms, and that, for this reason, truth could at least be partially analysed that way. One may object to this project of a piecemeal account of truth, however, by saying that we want to know what it is about the sentences that are true-in-mathematics that make them *true*. It is implausible to say that "true" means something different for different kinds of truth-bearers. That would imply that "What he said was true" is ambiguous, with different interpretation depending on the type of proposition. But contrary to "bank", this sentence just seems to mean one thing (cf. the argument in 1.5). Even if there were some argument for an ambiguity in "true", it does not seem that it should have anything to do with the nature of the proposition for which truth is considered. What, for instance, should we say about the sentence "It is true both that 1+1=2 and that snow is white"? Surely, this is not like the kind of pun exemplified by "The sky is blue and I am not". It will not do to say that "true" means "*F* or *G*", where these predicates are thought to analyse truth for the different types of propositions and say that the account makes "true" unambiguous (albeit disjunctive). On such a terminology, "bank" would be equally unambiguous (albeit disjunctive). If mathematical truth is to be explained in a certain way, then so must other kinds of truth, for there must be something in common in the different cases which is what makes both cases of truth.

There are also more topic-specific worries about epistemic theories for mathematical truth. Such theories operate crucially with the notion of *proof*, and hold that a statement is true iff there is a proof of it (cf. Dummett (1959), (1991: 333ff.), Prawitz (1998), Martin-Löf (1991)). These theories are called epistemic not so much in virtue of the truth-claim itself as in virtue of the

proponents' views about proofs as epistemically constrained. The most extreme, "finitist", versions hold that there is a proof for a statement only if someone has actually proved it. I should hasten to add that it is not always clear whether proponents of these theories have intended to analyse the notion of truth rather than merely making a claim about it. They have not been much occupied with questions of conceptual or explanatory priority, and those who merely wish to state a material equivalence, rather than a conceptual analysis or an explanatorily fundamental claim, need not bother with the argument from ambiguity above. But they are not proposing a "theory of truth", in the sense of 1.1.

This type of theory may seem to have good chances of capturing the extension of "true mathematical sentence", but we shall see, first, that it does so at a high price, and, secondly, that the claim is not plausibly taken as primitive in any case. The notion of proof is typically explained in terms of axiom, so that there is a proof of a statement iff it follows from the axioms. When considering which sentences or schemata are the axioms of a specific theory, one merely needs to point. More generally, one can say that the axioms of a theory are simply those formulae which function as axioms in the theory, i.e., those from which other sentences are to be derived. However, those who say that true mathematical sentences are those for which there is a proof do not intend their claim to be theory-relative, i.e., a characterization of being true-in-theory-*T*. They want to speak of the true mathematical sentences, period. But then, it would seem, they must say that a mathematical sentence is true (period) iff there is a proof for it, period. But looking back at the characterization of proofs, we see that this in turn requires that one can speak of *the axioms*, period. Thus, it follows that among the true mathematical propositions, some are axioms and others not, objectively and not relative to some system. But logicians typically compare logical systems in pragmatic terms, not in terms of right and wrong, in the sense that the right theory takes as axioms those propositions which really are axioms. This is thus a contentious consequence of these theories of mathematical truth, if the absolute notion of proof is to be considered in such close analogy with the formal notion. On the other hand, Dummett (1963: 201), for instance, denies this, recommending instead that "it should be [given] in terms of the inherently vague notion of an intuitively acceptable proof".

In either case, if the biconditional connecting truth and proof is taken as the correct theory of truth in the sense of 1.1, one can complain that, even if the biconditional is true, it is more reasonable to take this to be explained, or

derived, by the explanation of the notion of proof and a different truth-theory, rather than taking it to be primitive, i.e., underived. One can hold that the claim that something is true iff there is a proof of it informs one of a connection between truth and proof, but does not explain either notion, but rather to be a fact to be explained by the claims that do explain them, i.e., the correct theories of truth and proof, respectively. Is it not more natural to say that what explains the notion of proof is something like: a proof of a statement is a derivation thereof from (manifestly) true sentences? This does not conclusively show that such a theory together with a deflationary theory of truth gives the best overall theory. Conclusive evidence for such a claim is hard to come by. But our intuitions as to the naturalness and plausibility of taking various claims as primitive may well reflect the actual facts concerning which makes for the best overall account. In any case, the truth of a biconditional should not be taken as indicating that it is the correct theory.

1.3 DEFLATIONISM VS. ALTERNATIVES, PART 2: CORRESPONDENCE THEORIES OF TRUTH

It is commonly agreed that correspondence theories are the most serious rivals to deflationism. I here take "correspondence theory" to be one according to which truth should be explained in terms of representational notions, such as reference, picturing, satisfaction, or "expressing" (holding between, e.g., sentences and facts or predicates and properties). They may also, but need not, operate with notions of truth-*makers*, such as *facts* or *states of affairs*, which are thought to make truth-bearers true by their mere existence (or obtaining). Further, I take these theories to hold that truth is *explanatorily or conceptually dependent* on the notions in question. It is necessary to emphasise this dependence, because the mere truth of the biconditionals relating truth and, say, fact-expressing, is often granted by deflationists. The question is whether the equivalences are plausibly taken as primitive.

One point of the elaboration below is to characterise an important motivation for deflationism, which is its ability to do justice to the correspondence intuitions while avoiding the obstacles encountered by correspondence *theories*. Deflationary and correspondence theories thus share a certain feature, which is the view that for a certain truth-bearer to be true, it is sufficient that the world be a certain way, thinking beings need not in

addition stand in any cognitive or practical relation to the truth-bearer. (It should be said, however, that both theories are compatible with stark anti-realisms, even solipsism, since they may take what is the case in the world to be dependent on, or constructed by, minds. This combination seems never to have been endorsed, however, which may be of some metaphilosophical interest.) It seems reasonable that a general intuition supporting a correspondence theory is the obvious *relationality* of truth: whether, e.g., "Snow is white" is true depends on the way snow is (cf. Wright (1999: 208f.)). Therefore, it may seem, we must say what kind of thing in general needs to be in a certain way for a truth-bearer to be true. For instance, it may be held that a certain *state of affairs* (namely, the one that corresponds to the truth-bearer) must *obtain*, or that there is a *fact* of the appropriate kind (namely, one corresponding to the truth-bearer).

Deflationists typically accept the premise of this argument, but question the conclusion. *A fortiori*, it is held that deflationary theories can explain the dependence between truth and the way the world is just as well as correspondence theories. Briefly, if saying that "Snow is white" is true is just to say that snow is white, then, of course, whether the sentence is true depends on what the world is like. But deflationists do not only take the simplicity of their theory, in conjunction with its equal explanatory potency, to speak in its favour over correspondence theories. There are also additional arguments to the effect that the complexity of correspondence theories is idle in the stronger sense that, under scrutiny, they end up saying the same thing as deflationary theories. This is not always, or on all interpretations, the case, however. But where the additional complexity is not merely idle, it is often held to be positively implausible. As I will try to show here, these criticisms are related so as to present a dilemma.

Let us first separate three points that are taken to speak individually against correspondence theories. Though different varieties of these theories have been objected to because of features particular to their respective designs (that is why I refer to them in the plural), the following three each have at least one correspondence theory as its target:

(a)   that it commits itself to implausibly substantial accounts of facts or states of affairs,

(b)   that it only claims in an unnecessarily complex way what the deflationary theory says itself, and

(c)   that it requires that representational relations like reference and fact-picturing be reducible.

Though taking the points (a)-(c) to speak against correspondence theories is common among deflationists, there are few *loci classici* expressing deflationary criticisms of correspondence theories. The points are usually hastily passed by quite sloppily and laconically. I will refer to relevant passages as I go through the dilemmas.

Now, the dilemmas consist in the fact that correspondence theorists can avoid criticism by one point only by becoming vulnerable to criticism by another. In particular, avoiding (a) makes one vulnerable to (b) and avoiding (b) makes one vulnerable to (c). Having argued that this is so, it will emerge that the deflationist can agree with much of what correspondence theorists say, e.g., that true claims state facts, "correspond", in some sense, with them, and depend for their truth on their being facts. It will emerge, then, that the disagreement should be taken to concern only the question of which claim is most plausibly taken as primitive.

Point (a) is well-known from such pioneering deflationists as Strawson (1950: 139f., 153f.) and Quine (1960: 246ff.) and (1987: 213). According to them, facts, states of affairs, etc., should not be unduly reified, that is, should not be explained, or taken to exist, in any substantial way. Of course, something positive must be said about these notions, for it is highly implausible to regard all statements involving them as false or mere nonsense. The typical deflationist answer, given by Ramsey (1927, p. 158f.), Strawson (1950: 136ff.) and Quine (1987:213), is to give a *deflationary theory of facts and states of affairs*. For instance, Ramsey says:

> We can, if we like, say that [the proposition aRb] is true if there exists a corresponding fact that a has R to b, but this is essentially not an analysis but a periphrasis, for «The fact that a has R to b exists» is no different from «a has R to b».

According to such a theory, what explains the notion of fact is rather that:

(i)   "The fact that *p* exists" is trivially equivalent with "*p*",

or

(ii)   "That *p* is a fact" is trivially equivalent with "*p*".

It will now be argued that if the correspondence theorist is not to give an implausible account of facts, as charged by (a), he must treat them as understood according to the equivalences above, but that if he does, then he will be subject to criticism on behalf of point (b) (Wright (1999: 218f.) and Blackburn (1984: 224ff.) both make similar points). Let us begin to look at two simple variants of correspondence theory:

> (TA1) A sentence is true iff it asserts that a fact exists, which does in fact exist.

> (TA2) A sentence is true iff it asserts a fact.

If we assume (i), a deflationary theory of truth, and every instance of " '*p*' asserts that *p*", we can derive (TA1). Likewise, assuming (ii), a deflationary theory of truth, and every instance of " '*p*' asserts that *p*", we can derive (TA2). But this shows that the plausibility, indeed the truth, of (TA1) and (TA2) in no way gives reason to think that there is some explanatory dependence between truth and facts. In other words, the truth of (TA1) and (TA2) is compatible with the claim that truth should not be explained in terms of fact (or, of course, *vice versa*).[1] This also shows that the corre-spondence theorist, insofar as he wants to avoid point (a), faces point (b), according to which he has given an unnecessarily complex, idle rephrasing of the deflationary claim.[2]

---

[1] David (1994: 2.5) argues, quite differently, that if one holds that propositions are the primary truth-bearers, then one would be a correspondence theorist, since true proposi-tions *are* facts (which in turn *are* obtaining states of affairs). This reasoning illegitimately precludes by *fiat* the view that combines deflationary theories of both truth and facts with the view that facts are true propositions. Again, this latter claim does not entail any ex-planatory dependence between truth and facts (in either direction).

[2] Dodd (2000: 5) argues that because correspondence theories necessarily hold that for every true proposition there is something such that its existence entails that the proposition is true, they are *ipso facto* substantial, and point (b) fails. The necessity here seems, first, to stem from Dodd's rather optional definition of "correspondence theory". But secondly, if this correspondence claim can be shown to follow from a deflationary theory of propositions and (i), then point (b) will indeed pose a threat to the corre-spondence theorist (in Dodd's sense). To see that it does follow, suppose that the proposi-tion that *p* is true. By deflationism, it follows that *p*. By (i), it follows that the fact that *p* exists. From this, it follows, i.e., it is *entailed*, that the proposition that *p* is true (just reverse the two previous steps). Thus, if the proposition that *p* is true, then the fact that *p*

An analogous situation faces the correspondence theorist who prefers to speak of states of affairs. If this notion is explained by the equivalence between sentences "The state of affairs in which *p* obtains" and the corresponding "*p*", then such a theory will also be demonstrably true, but also unnecessarily complex. They will also be misleading by suggesting an explanatory dependence of truth upon states of affairs. (If the theory says that there is such dependence, then it is not true, of course, but this is not normally explicitly claimed, but implied by the pride of place the theorist gives to the biconditional.)

Thus, (TA1) and (TA2) hold because of the equivalence between sentences "That *p* is true", "That *p* is a fact", and "*p*". This equivalence cannot be denied. What the correspondence theorist must deny is that the notions of truth and fact are explained by these equivalences, for then, he is a deflationist regarding both truth and facts. It seems the correspondence theorist has no choice but to commit himself to a substantial notion of facts or states of affairs.

We can at this point refine another common argument against correspondence theories, which holds that they cannot explain those equivalences that the deflationist takes to be primitive (cf. Horwich (1998a: 11f.)). Though an important argument, it needs to be refined, for given, e.g., (i), which is a deflationary theory of fact-existence, and (TA1) or, alternatively, (ii) and (TA2), these equivalences can easily be inferred. The argument should therefore be instead that the correspondence theory cannot explain the deflationist's equivalences unless it takes the notion in the *analysans* (e.g., fact) to be explained by a corresponding equivalence, which is to give a deflationary of that notion. Thus, the theory will be either explanatorily inadequate or unnecessarily complex. Though the examples of (TA1) and (TA2) is insufficient to prove this general point, they may indicate the general point that any correspondence theory must at some point give a deflationary theory of some notion appealed to in the *analysans* in order to derive the equivalences.[3]

---

exists, and that the fact that *p* exists entails that the proposition that *p* is true. By existential generalisation, if the proposition that *p* is true, there is something that exists which is such that it is entailed by its existence (or: by the proposition that it exists) that the proposition that *p* is true. Since the proposition that *p* was arbitrarily chosen, we may derive the correspondence claim.

[3] Horwich does note, however, that some correspondence theories he considers will end up with a schema in the characterisation of fact (1998a: 107).

Let us now look at how points (b) and (c) are related. Note that a substantial account of fact or state of affairs, together with a theory such as (TA1) or (TA2) will not suffice to determine whether a truth-bearer is true given what facts exist or what states of affairs obtain. One would also need to know which fact or state of affairs matters to the truth of a particular truth-bearer. In other words, one needs to know the nature of the correspondence between truth-makers and truth-bearers. It will not do to say that what would make the proposition that snow is white true is the existence of the fact that snow is white, and so on. According to such an explanation, it is the schema "The fact that p corresponds to the proposition that p" which explains the relation. No such manoeuvre is available to correspondence theorists because they do not regard correspondence to be something to be stipulated so as to make the truth-analysis come out true. They regard the relation of correspondence as substantial and one that is a non-trivial matter to explain. This schema must thus rather be regarded as a criterion of correctness upon the elucidation of correspondence, in the sense that such an elucidation must make every instance of the schema come out true. This means that in order to avoid point (b), the correspondence theorist needs also to commit himself to the possibility of elucidating a substantial representational relation of correspondence, which is the charge of point (c).

Classical examples of explanations of such a correspondence relation are those of Russell (1912), Wittgenstein (1922) and Austin (1950). I cannot here enter into a discussion of these theories here, but to get a flavour of the problem, consider what is involved in explaining what it would be for the sentence "Something is red" to correspond in some substantial way to the fact that something is red, and, secondly, what substantial thing the fact that something is red is supposed to be. It can hardly be: *something red*! But what, then? Negative and disjunctive facts have also been considered a problem for the substantialist about facts (whereas they do not seem to pose any problem for the deflationist about facts). Whatever these facts are, further, how does a sentence or proposition correspond with it when and only when it is true, other than in the obvious way granted by the deflationist?

It is this type of criticism is that summed up in point (c), and we have thus come to a close in our critical examination of correspondence theories. First, it was argued that substantial accounts of facts are implausible. Next, we saw that by avoiding this criticism by giving a deflationary account of facts, the truth-analysis became indistinguishable from deflationary theories, albeit unnecessarily complex and confusing in its implication that there be some ex-

planatory asymmetry between truth and fact. To avoid that criticism, the correspondence theorist needs to invoke substantial facts and some substantial correspondence-relation between them and truth-bearers.

Now, finally, I should like to give an explanation congenial to a deflationist of the very idea that we should appeal to facts or states of affairs in explaining truth. It cannot be the fact that whether something is true depends on how the world is, for that fact motivates deflationism as much as a correspondence theory. An better explanation, I believe, is this: if sentences of the form "the fact that p exists" and "the state of affairs in which p obtains" are equivalent to the corresponding sentence "p", then these locutions increase the expressive power of a language in precisely the same way that "true" does. Now, what is wanted in the *analysans* of (TA1), for example, is a sentence having all instances of "x says that p and p" as consequences, viz., some kind of generalisation, but one that does not contain "true". More generally, what is needed in the *analysans* is a sentence which covers an infinite number of sentences which are the instances of a schema with schematic sentence-letters. Given the equivalence noted above, the locutions of "fact" or "state of affairs" can do precisely this, and this is why truth-analyses like (TA1) are true, indeed necessarily true. The mistake in appealing to facts, etc., is that of conflating necessary equivalence with explanation. The *analysantia* mentioning facts are indeed equivalent to the truth-ascriptions, but do not explain them, because both halves of the analyses should be explained by recourse to deflationary schemata. These, in turn, show why the analysis is true. It is thus not correct to say that a deflationist must deny that truth is correspondence to facts or states of affairs in order to be a deflationist. But the "is" here must not be understood as expressing an explanatory relation, but only a necessary truth. To explain such equivalences is a criterion of exhaustiveness of the truth-theory, since it is indeed a fact that a sentence is true iff it says that a state of affairs obtains, which really obtains, etc. Since a simpler overall account is had by taking the simple equivalences concerning truth, fact, and state of affairs, respectively, as primitive, and the claims connecting these notions as derived, the deflationary theories are vindicated.

In view of the above reasoning, we can now also better understand the oft-repeated argument against correspondence theories, that judging whether something is true would involve an unintelligible comparison between a truth-bearer and a truth-maker, so that, e.g., in order to judge whether the belief of mine that snow is white is true, I would have to examine a certain

fact and then go on to see whether the appropriate relation between this fact and the belief holds.[4] Now, if "fact" is understood in a substantial way, this argument is perhaps cogent – the comparison model does seem to give a rather unnatural account of ordinary cognitive processing; one, namely, according to which there is a process from fact-perception, through fact-thought comparison, to truth-judgment over the thought (*pace* Schlick (1935: 65ff.)). But on a deflationary theory of facts and truth, it is not, for there, judging that the fact, e.g., that snow is white, exists and judging that the thought that snow is white is true come to the same thing, namely, judging that snow is white. The fact-mentioning account of our judgmental activities would then not be perverse other than in its wordiness.[5] The deflationist can thus agree that truth is correspondence to facts (though not that this explains truth) without being vulnerable to this charge.

1.4 Six Central Theses

Due to the variety of deflationary claims, I have found it convenient to list a number of theses, with ensuing comments, that have been more or less strongly associated with the term "deflationism" here at the outset, as points of reference for the rest of the book. It should be kept in mind that it is doubtful whether any self-proclaimed deflationist has endorsed all of the theses, yet all of them have endorsed at least one of the theses, and, moreover, each thesis has been endorsed by at least one of them. The term "deflationism" with lower-case "d" will be used throughout this book to designate, quite vaguely, views which exhibit the kind of spirit motivating these theses. The author's preferred stipulation of "Deflationism" (with capital "D"), however, is thesis (I):

---

[4] Various formulations of this argument are found in Neurath (1934), Hempel (1935a: 50f.) and (1935b), Davidson (1986: 307) and Williams (1977).
[5] This may have been the intent of Ayer (1935).

(I)   Claiming the equivalence (of some yet unspecified kind) between some (or all) sentences of the form

      (i)   It is true that p,
      (ii) (The proposition) that p is true
      (iii) (The sentence) "p" is true

and the corresponding sentence "p" is sufficient to give an exhaustive account of truth (the criteria of exhaustiveness will be discussed in 1.5).

This claim is one way of generalising the particular claim that for the proposition that snow is white to be true is just for snow to be white.[6] Candidates for the equivalence relation referred to in (I) may be notions like synonymy, co-assertibility, interderivability, intersubstitutability, etc. It is usually suspected that the spelling out of notions like these is a difficult and serious matter, because deflationism requires that we are not, in spelling them out, relying on some non-deflationary concept of truth. We will assume throughout that this can be done (cf. 3.4).

---

[6] Some basic terminology: (i)-(iii) are called *schemata* or *schemas*, and sentences of those *forms* (also called their *instances*), e.g., "It is true that snow is white", are had by substituting for "p" an English declarative sentence. The *corresponding sentence* "p" in thesis (I) is simply the sentence which substitutes the "p" in (i)-(iii). The "p", as occurring in the schemata is called a *schematic sentence-letter*. Further, sentence schema (i) only touches the "truth-*operator*", "It is true that", while the other two contain the expression "is true" – the truth-*predicate*.

    Some authors confusingly use the expression "p is true" (notably Ramsey (1927: 158) and Carnap (1942: Def. 17-1)) when discussing schemata relevant for deflationism. However, unless "p" here is taken either as a first-order variable (ranging over sentences or propositions) or as a name of a sentence or a proposition, but is interpreted rather as a schematic sentence-letter (as in (1)-(3) above), then the instances of "p is true" are ill-formed. One cannot *apply* a predicate to a sentence, viz. write "is F" after it (though to a description or a name of a sentence one can), though one can *ascribe* a predicate to a sentence, viz., say of a sentence that it is F. Further, one cannot say of a name or description of a sentence that it is true, though this is not a matter of grammar. A cognate mistake is that of speaking of a sentence s being true and then using the expression "that s" as a grammatical object for propositional attitude verbs or the truth-predicate. The appropriate grammatical object would have to be referred to using the concatenation-function (see 3.1).

(II)   Some claim about one (or all) of the following schemata suffices to give an exhaustive account of truth:

> (ES) It is true that *p* if and only if (iff) *p*
> (PS) (The proposition) that *p* is true iff *p*
> (DS) (The sentence) "*p*" is true iff *p*.

Concerning thesis (II), the schema (ES) (spelled out "Equivalence Schema", after Dummett (1973: 445) and (1978: xx)) is the schema touching the truth-operator. The schema (DS), furthermore, is spelled out "Disquotation Schema" (after Quine (1970: 12)) because a sentence in which "is true" is applied to a quote-name of a sentence just says what the quoted sentence says. Thus, for such a truth-ascription, just erase "is true" and the quote-marks and you get an equivalent sentence. The schema (DS) is favoured by those who take sentences to be primary truth-bearers. Finally, (PS), the "Propositional Schema" is focussed at by those who take truth to apply primarily to propositions. By analogy with the Disquotation Schema, (PS) may be called a "denominalisation schema", in that the expression "is true" and the "that", which forms the nominalization (a "that"-clause) disappears in the equivalent sentence. Note also that sentence type (i) in (I) stands to (ES) as (ii) stands to (PS) and (iii) to (DS). Which of these schemata one takes to be primary reflects one's views about what one takes the primary truth-bearers to be. If one takes sentences to be primary truth-bearers, for example, one will regard sentences of type (iii), or the schema (DS) as primary. This question of primacy will be addressed in Chapter 4.

    It is easily seen that (I) and (II) are closely related. In fact, if we endorse (II) by taking the deflationary theory to say that all instances of (ES) are *true*, and hold that this gives an exhaustive account of truth, then we automatically subscribe to (I), since this simply is a way of claiming a certain kind of equivalence between sentences "It is true that *p*" and corresponding "*p*". There are of course many other ways of assimilating (II) to (I). However, for reasons that will be descried in Chapter 3, there is a point in distinguishing them.

(III)   No Truth Analysis of the form

(TA) For all *x*, *x* is true iff …*x*…

(where "…*x*…" represents some sentence with only the occurrence(s) of "*x*" free) yields a correct theory of truth.

This is a thesis common to almost all deflationists.[7] Of course, this does not mean that no claim of the (TA)-form can be true, but only that if it is, then the truth-equivalences appealed to by the deflationist must be able to explain it. This is related to, but not identical with the claim that "truth has no explanatory function", and that "truth has no underlying nature to describe", which will be discussed below.

(IV)   The correct theory about truth (itself) is one which only describes the semantic properties (in a restricted sense) of the word "true". (Alternatively: nothing can or need be said about truth proper, but only about the semantic functioning of the expression "true".)

It has often been stated in the literature that deflationary theories tend to take a more "linguistic" form than other truth-theories (cf. Kirkham (1992: 30f.), McGinn (2002), Devitt (2002)). Thesis (IV) is a preliminary attempt to clarify this idea. However, it would seem that without a certain restriction on the notion of "meaning" or "semantic properties", one could endorse (IV) without being a deflationist. Perhaps it would then also be compatible with *any* (non-sceptical) truth-theory, and thus trivial. In order to see that (IV) can be made compatible with a correspondence theory, for instance, one could consider the claim that, according to the correct description of "true", it is used to pick out a substantial property, e.g., a relational property which holds between linguistic or mental entities and facts or objects in the world. The deflationist wanting to endorse (IV) thus has to restrict the notion of "semantic properties", so as to exclude an interpretation according to which the description of "true" above would be a semantic description.

---

[7] Exceptions are the deflationary theories discussed in Baldwin (1989), David (1994) and van Inwagen (2002). Here we find truth-analyses of the (TA)-form, but all of them contain a propositional quantifier in the *analysans*, which is telling, since that seems to be the only way for a deflationist to reject (III). This kind of truth-analysis is discussed in 3.3.

But there still seems to be something right about (IV), when seen in the light of theses (I) and (II). There, the semantic explanation is purely *intralinguistic*, as opposed to the semantic explanation appealing to property-reference (the reading that made (IV) compatible with correspondence theories). It seems that we have to give it a reading on which it follows from (I) and (II) (separately). This could be done, it might be thought, since in those theses, the claim that is supposed to give an exhaustive account of truth is one where only the word "true" (or, to be precise, a form of sentence containing the word) is *mentioned*. By contrast, in analyses of the (TA)-form, the word is *used*. As we have just seen, a correspondence theory can be reformulated so that "true" is not used, but only mentioned, but a reverse move for deflationism seems not to be possible.[8] This, then, could be the correct elucidation of (IV): the correct theory of truth only mentions "true" and *cannot* be reformulated so that it uses it.

The question whether we should say that a theory about the word "true" is "about truth" or that it is not, adding that nothing can be said about truth we may leave open. The predicament is not unique: should we say, for example, that we have a theory of existence when we have a complete true theory of the existential quantifier and the existence-predicate or only that we have a theory of these linguistic expressions? In this case, most people agree that no further theory of "existence itself" could be had. In the case of truth, the deflationist who endorses (IV) believes something similar. A related point is the fact that theories of the (TA)-form can be reformulated by giving claims of the form "Truth is …" or "To be true is …", to be filled in, e.g., by "correspondence to the facts" or "to correspond to the facts", respectively. But nothing of this kind can be done with a deflationary theory on the lines of theses (I) or (II) (cf. David (1994: 65ff.)). In any case, it is incorrect to speak as if a deflationist could agree with a substantial theory of truth itself, since deflationism only deals with the *word* "true". If anything is essential to deflationism, it is the denial of such a proposal. The questions about the linguistic character of deflationism will be discussed at length in Chapter 3, the conclusion of which is precisely thesis (IV) under the proposed interpretation.

---

[8] The exceptions are deflationary theories of the (TA)-form that use propositional quantifiers, which will be discussed in 3.3.

(V)   There is no property of truth (or: truth is not a property).

Many have explicitly held thesis (V) to be central to, and even defining, deflationism. This seems to be a result of regarding it as a consequence of some (or perhaps all) of theses (I)-(IV) and (VI). However, it is difficult to see why (V) should follow from any of these without some further premise about properties. Further, such prominent deflationists as Paul Horwich (1998a: 37) deny (V) and Hartry Field seems to take it as independent of deflationism (1994: 265n.19). It is therefore difficult to see why it should be associated with deflationism at all.

It is more commonly said that deflationism holds that there is no *substantial* property of truth. The problem is then to explain this metaphor of "substance". Perhaps most philosophers discussing this issue mean, by "substantial property", one analogous to *heat*, which is substantial in virtue of being reducible to an unobvious, conceptually distinct property (namely, mean molecular kinetic energy). Horwich says that truth has no "underlying nature" to be revealed (1998a: 2), which is probably intended to mean precisely that it is unobvious, i.e., not *a priori* knowable, but "hidden".

There may of course be further characterizations of "substantial property", but it is safe to say that whatever they may be, it should follow directly from the claim that the equivalences exhaust the notion of truth, that truth is not substantial in the given sense. This holds trivially on the above elucidation, for if the equivalences are *exhaustive* of truth, there cannot be any further identification such as a physicalistic reduction. It is because this type of denial follows directly from the exhaustion claim that some deflationists have said that "deflationism holds that there is no substantial property of truth", which, in turn, has encouraged the *identification* of deflationism with this negative claim. On such a definition, however, deflationism is compatible with the claim that the equivalences have nothing to do with truth, that truth is inexplicable, and that it has no expressive function. Clearly, this is a non-standard definition which goes counter to the spirit of every professed deflationist.

The regrettably unexplained notion of "substantial property", finally, has figured in a crucial way in three debates in particular: the ones prompted by Crispin Wright's (1992: Ch. 1) and Paul Boghossian's (1990) arguments against deflationism (Horwich replies to both in his (1998a: 142ff.)), and the one over whether the fact that adding (DS) as an axiom-schema yields a non-conservative extension of a theory entails that truth is a substantial property

(and thus, allegedly, that deflationism is false) (Shapiro (1998)). But when the claim that truth is not "substantial" is substituted for the exhaustion claim, the arguments lose their bite, plausibly because the obscurity of "substantial" cannot then be polemically exploited. For instance, it is unintelligible why the fact that (DS) yields a non-conservative extension should be an objection to the exhaustion claim.

(VI)   The only important point (in some sense) of having "true" in our language is to increase its expressive power.

This thesis is an attempt to make plausible the combined view that truth is just as simple as it is according to theses (I) or (II), but that there is still an important point in truth-talk. It could accordingly be taken as a reply to the objection that if (I) or (II) were true, then there would be no point in having a truth-predicate in one's language. The thesis is not sufficiently specified, however, in that the notion of "important" is not made clear. The idea is to exclude such functions of "true" as serving stylistic or ornamental purposes, and include the theoretical uses – those uses we are interested in when doing philosophy.

The idea about the expressive function can be explained thus: there is a certain point in using the word "true" in *some* sentences, while in others, it is redundant. For instance, in "It is true that snow is white", the use is redundant, because this sentence only says what "Snow is white" says, and thus, "true" is eliminable. However, in

(1)   Everything he said is true,

we cannot give an equivalent sentence lacking "true", which shows that in (1), the use of "is true" is irredundant. These ineliminable occurrences of "true" are examples of what the expressive use of "true" comes to.[9] For note that all sentences of the form "If he said that *p*, then *p*" follow from (1). What makes "true" special is that if we lacked "true" (and all equivalent expressions, like "holds", "is a fact", etc.), we could not form a sentence with this inferential property, i.e., which implied all sentences of the form "If he said that *p*, then *p*". Likewise, such blind ascriptions as "What Percy says is true" has all conjunctions of instances of the schema "If what Percy says is

---

[9] Cf. especially Quine (1970: 11f.), Leeds (1978: 121f.), Soames (1990: 230).

that *p*, then *p*" as consequences. We will say that "true" increases the *expressive power* of the language, and define this as follows. Let $I(s, L)$ be the class of all and only the sentences of *L* which are analytic entailments of *s*. We can at this stage let paradigmatic examples characterize the notion of analytic entailment and say that, e.g., "*x* is a mother" analytically entails "*x* is a woman" and all conjunctions analytically entail their conjuncts. Now, we can define expressive strengthening thus:

(EP)   An expression *e* increases the expressive power of a language *L* $\equiv_{df} (\exists s_1 \in L \cup \{e\})(\forall s_2 \in L) (I(s_1, L) \neq I(s_2, L))$

Note that, e.g., an expression like "tiger" would not increase a language's expressive power in this sense, since the expression is required to be such that sentences which do *not* contain it are implied by sentences that contain it. For example, instances of "If he said that *p*, then *p*" obviously need not contain "true", though they are implied by (1), which does. The word "tiger" would increase the expressive power of a language in the sense that one can say things with it, e.g., that there are tigers, which could not be said without it or a synonymous expression. But it would not increase the expressive power of a language in the sense given by (EP). For the definition to distinguish between "true" and "tiger" in the intended way, however, the sentences quantified over may not be infinite conjunctions or disjunctions, for then, there would be a sentence entailing every instance of "If he said that *p*, then *p*", namely, the infinite conjunction of these instances. Secondly, depending on fact about "analytical entailment", we might need to require that the classes be non-empty, if "tiger" is not to count as increasing expressive power. This is because there might be sentences containing "tiger" with no analytic entailments in the language lacking "tiger", e.g., "Tigers are tigers". If so, no sentence in the sublanguage would have the same analytic entailments, since every such sentence would entail itself, and "tiger" would satisfy the *definiens* of (EP).[10] This is not needed, though, if all analytic sentences are analytic entailments of any sentence, for then the relevant class would be identical to that of any analytic sentence. It is also not needed if "Tigers are tigers" entails, e.g., "Some things are the way they are", or some such, and every other sentence with "tiger" entails some sentence in the sublanguage.

---

[10] I owe Peter Pagin for this second observation.

The fact that the truth-predicate seems to be expressive in this sense, and the difference between it and, e.g., "tiger", is probably the reason why many tend to say that it is a "merely logical" predicate – it has no empirical content of its own, and is thus more like a sentential connective than "is a tiger". One might say that "true" is a kind of "content-function" in that a context "That *p* is true" transparently takes the (possibly empirical) content of the embedded sentence to itself.

However, it is usually thought that even if actual truth-predicates increased the (natural) languages into which they were introduced, they do not increase the expressive power of any language. In particular, if a language contains something like propositional quantification, then it would not need "true" (cf. Soames (1999: 256, n. 4)). For instance, we would not in such a language have to assert the sentence (1) in order to say something which implies all instances of "If he said that *p*, then *p*", for this job would be done by "For all *p*, if he said that *p*, then *p*". At least, this is how such quantifications into sentence position are intended to behave inferentially, though there is disagreement as to whether anyone has adequately defined them. The deflationist's story about this fact is typically as follows: introducing propositional quantification would involve the introduction of wholly new ways of forming sentences. For some reason, natural languages are conservative when it comes to certain grammatical innovations, and therefore, other ways of enabling the expressive power, which do not introduce new grammatical categories, have evolved.[11] (Such grammatical innovations in natural language, however, is by no means unthinkable: the so-called prosententialists describe a variant of English with precisely the required addition (as we will see in 2.5), as does Prior (1971: 37).) So, instead of introducing a new grammatical category, an expression of an old grammatical type is introduced, but with a new kind of *semantic* property, in order to accomplish the expressive strengthening.

That semantic property is plausibly taken to be the one described in (I), sentence-form (ii). Given (I), the construction of a sentence with the power to imply all instances of "If he said that *p*, then *p*" is simple: first note that sentences of this form are equivalent, by (I), to those of the form "If he said that *p*, then that *p* is true". Since this, in turn, is a sentence of the form "If he said *x*, then *x* is true", one can form a universally quantified sentence with all the instances thereof as consequences, namely (1).

---

[11] See, e.g., Grover *et al.* (1975: 99f.), Forbes (1986: 36), Horwich (1998a: 32).

Finally, it is important to observe the exact wording of (VI). What marks deflationists off from their opponents as regards (VI) is not that only deflationists think "true" increases the expressive strength of our language – almost everyone agrees with this. Rather, it is that a deflationist will hold this to be the *only* function of the locution, while non-deflationists tend to think that "true" has other functions as well. (In claiming that this is the only function, I am neglecting ornamental or pragmatic uses.) Normally, the deflationary view is taken to be that "true" has no *explanatory* function, as opposed to what non-deflationists would say. But it is rather problematic what it is for a word to be explanatory. Clearly, it must not mean that no facts are to be explained by saying that certain things are true, for it can obviously be correct to explain, e.g., someone's success by saying that his beliefs were true. An example might clarify the intent of (VI). When epistemologists identify knowledge with justified true belief, the use of "true" should be seen merely as a device of covering in a finite statement all instances of "X knows that p iff X has a justified belief that p and p". Thus, according to deflationism, a substantial reduction of truth should not be expected, and is not needed, in order to have a complete theory of knowledge. It seems that the crucial claim that deflationists must reject is that "true" is like natural kind terms in referring to some property such that describing the nature of that property will be part of an explanation to the putative facts about things with the property. For instance, the description of the nature of the property referred to by "hot" partakes in explanations to why hot things melt, or evaporate, or expand, etc. That should be the kind of "explanatory function" of "true" denied by the deflationist. In Horwich's words, truth is not "a characteristic whose underlying nature will account for its relations to other ingredients of reality" (1998a: 2).

1.5 CONSTRAINTS ON DEFLATIONARY THEORIES

In this section, I will list six adequacy constraints upon a deflationary theory of truth. The first two are specific to deflationism, the third is specific to truth-theories in general, and the last three are among such constraints that any theory must meet, but which are mentioned here because they may be seen as particularly problematic for deflationism.

The first two constraints are such that, if a deflationary meets them, it gives, by my definition, an *exhaustive account* of truth, and thus, they help in specifying the sense of theses (I) and (II), where this notion occurs crucially. The first, which we might call the *explanatory constraint*, is that the theory must "explain all facts concerning truth". The alleged facts mentioned have included the fact that true beliefs facilitate successful behaviour, that meaning is truth-conditions, that it is the truth of theories that accounts for their empirical success, that true conjunctions have true conjuncts, and many more.

Paul Horwich (1998a) has been prominent in showing such alleged facts to be either accounted for by (his version of) deflationism, or actually not facts at all. One important point in connection with our present concern is to see the way in which such facts are thought to be "explained". Horwich does this by using instances of (his variant of) the truth-explaining schemata, together with other plausible premises taken from other areas, e.g., concerning action, meaning, empirical theories, etc., in order then to *infer* the facts. So, the method is to show how the facts follow from his theory of truth together with other facts. Let us look at an example of how Horwich attempts to explain the fact that true beliefs facilitate successful behaviour.[12] Horwich here asks us to

> [c]onsider […] those of a person's beliefs of the form
>
> > <If I perform action *A* then state of affairs *S* will be realized>.
>
> The psychological role of such beliefs is to motivate the performance of *A* when *S* is desired. When this process takes place, and if the beliefs involved are true, then the desired result will in fact obtain. In other words, if I have belief (1) and desire *S*, then I will do *A*. But if my belief is true, then, given merely the [instances of (PS)], it follows that if I do *A* then *S* will be realized. Therefore, by modus ponens, *S* will be realized; I will get what I wanted. (1998: 44)

It is not essential that this type of account yields, in some sense, a *complete* explanation of the fact, but rather that it can be shown that no more needs to be said about truth in order to completely explain it, whatever a complete explanation is supposed to be. Success in showing this, furthermore, simply

---

[12] This kind of explanation of truth and successful behaviour originally derives from Williams (1986: 232f.).

consists in showing that the alleged facts follow from the theory of truth and claims about other matters. Perhaps a complete explanation will require more, but Horwich's view (1998: 45), with which I agree, is rather that the full explanation of the fact will not require more from the theory of truth than (some generalisation of) the equivalences. Of course, this view is quite contentious – in fact, it may well be said to be *the* controversy surrounding deflationism – and thus one which cannot be judged at this stage.

The second constraint, to be called the *inferential constraint*, also taken as common ground in the debate, is that the theory justify valid inferences which depend for their validity on occurrences of "true". For instance, the argument

(A)   It is true that snow is white.
       If snow is white, then snow is not colourless.
       _____

       Therefore, snow is not colourless.

needs to be supplemented with a further premise, and this premise must follow from the theory of truth, if the theory is adequate. To see that this requirement need not be that easy to satisfy, consider

(B)   Nothing Descartes believed is true.
       Descartes believed that he existed.
       _____

       Therefore, Descartes did not exist.

In such complex cases, we need to show the arguments to be valid given the theory of truth and predicate logic, together with specifications of how to deal with quantifications over truth-bearers. To clarify, an argument that depends for its validity on an occurrence of "true" is one where one could replace "true" with another word to get an invalid argument. For instance, the argument "Snow is white and something is true; therefore, something is true" does not depend for its validity on the meaning of "true", whereas (A) and (B) do. To "justify" these arguments means to give a theory (the generalisation of the deflationist equivalences), which, together with the premises at hand *logically* implies the conclusion.

One might perhaps say that the inferential constraint is a "subconstraint" of the explanatory one, in that meeting the latter implies meeting the former.

Thus, one might call it a (trivial) *fact* that if everything Descartes believed is false, and he believed that he exists, then he did not exist. This fact, furthermore, is to be explained along the lines envisaged by Horwich, by inferring the fact from the theory. The difference between the constraints is that the inferential constraint does not involve premises from other areas, while the explanatory one does. That, then, is the explication of the notion of giving an exhaustive account of truth. The deflationary claim is that, since the (preferred) claim about the equivalences is sufficient for giving, in the explained sense, an exhaustive account of truth, it is the right theory.

The third constraint is that a theory of truth must accord with the actual concept of truth and not give a stipulative definition of "true" that gives it a semantic interpretation different from that which ordinary speakers associate with it. Let us call it the *constraint of descriptivity*. How this should be spelled out and whether it should be met by a correct truth-theory are quite controversial matters, so I will here clarify and argue for the constraint at some length. The basic idea can be put as a truism: if we want a theory of what it is to be true, then we want a theory of what it is to be true, given what we ordinarily mean by "true" and nothing else, just as if we want a theory of what it is to be red, we want that and not a theory of what it is to be red in the sense of "red" as orange, round or anything else distinct from red. I am not excluding the possibility that the truth-theory describe *more* than the semantic character of "true", e.g., that it gives an *a posteriori* naturalistic reduction on the lines of "Water = H_2O". (That the correct truth-theory will only give a semantic characterization will be argued later on in this book, however.) The present constraint is only that the truth-theory not *violate the meaning* of "true". For instance, it may not entail that a sentence containing "true" is necessary, if the ordinary meaning of "true" is such as to determine it as not necessary (and similarly for other properties of sentences that are determined by meaning). That is, shared intuitions that a sentence is has a property, where this is determined by meaning (perhaps necessity, aprioricity, etc.) should tell against any theory which entails that it does not, and *vice versa*. But the constraint does not rule out a theory on the grounds that it entails that a sentence containing "true" has a property, where this is not determined by the (ordinary) meaning of "true". The sentence " 'Snow is white' is true" is true, but this is not determined by the meaning of "true", only partly so.

The analogy with "red" may now be disputed in two ways, so as to cast doubt on the constraint of descriptivity. First, one may say that "true" is

ambiguous, and that there is therefore a point in disambiguating by stipulatively defining distinct predicates that each capture one of the many interpretations. This is often the case, of course, putative examples being "meaning", "cause", "rational", etc. But "true", it seems, is not ambiguous. If it were, then a sentence containing "true" could be shown, given a context, to have different meaning-related properties, such as having a certain truth-value, a degree of confirmation or being assertible by someone, depending on different reasonable interpretations of "true" (cf. Srzednicki (1966: 387)). This obviously holds for, e.g., "bank", but, it seems, not for "true". At least I have never come across such a sentence, though I invite the reader to try for herself.

This fact does not, however, rule out the possibility that speakers use "true" unambiguously within their idiolect, but mean different things from one another, and that it is such differences that surfaces in the differences in philosophical opinion. The concurrence in speakers' overall use of "true" could be taken as a consequence of the fact that every speakers' individual interpretation has same consequences in most normal contexts, but in contexts pertaining to philosophical theories of truth, disagreement arises in virtue of semantic differences among speakers.[13] One could here appeal to the empirical investigations of ordinary speakers' responses to philosophical questions, prepared and interpreted by Arne Næss (1938) and others of the Oslo School. These questionnaires included statements as to the meaning of "true", e.g., that it means "agreement with reality", etc. Unsurprisingly, different speakers responded in different ways, as do philosophers. This only shows that people, whether philosophers or not, disagree on truth. The relevant question is whether these differences should be taken to reflect differences in semantic interpretation. It seems that, in the absence of additional arguments to the contrary, it should not. One reason is that people often give confused descriptions of the meaning of the words in their vocabulary, i.e., descriptions which are definitely not reflections of their semantic interpretation. For instance, people assent to sentences like "Truth is essentially an invention by the ruling class", "Truth is essentially the will of God", etc. The fact that such implausible claims are common, and that also more reasonable claims about meaning and essence are often in contradiction with facts about what the speakers assent to in more mundane situations, shows that claims about the meaning or essence of truth are not reliable

---

[13] Peter Pagin raised this point to me.

indicators of semantic interpretation. So, even where people assent to things that are logically or conceptually incompatible, this is not conclusive evidence for a difference in semantic interpretation. Further, of course, even where the difference is semantic, it must be shown that the difference pertains to "true", rather than some other word. Therefore, the burden of proof is still on those who claim that these differences must be explained by semantic differences of "true". Naturally, people can also be entirely clear over their semantic interpretation of a word while having different factual opinions about the "thing" it stands for. The constraint of descriptivity, again, does not exclude this possibility for "true". The above is only to respond to the idea that a truth-theory may not be in accordance with the actual meaning of "true", since it has more than one meaning.

Some possible confusions concerning this last proposal should be cleared away. If, by "difference in interpretation", one means difference in certain neural mechanisms responsible for the semantic contribution to overall speech dispositions, then it is doubtful whether this is a difference in semantic interpretation at all, the latter being more plausibly taken as a multiply realizable state. Concerning speakers' speech dispositions that do not concur, the question is whether one should appeal to mechanisms underlying semantic interpretation, or other cognitive (or emotive, etc.) neural states, and the above argument applies. One may also mean, by "difference in interpretation", that speakers internalize different (propositional or verbal) definitions. The difference in philosophical opinion may then simply be a direct consequence of which definition is internalized. But it is rather contentious to assume that semantic interpretation is effected by internal definitions at all, whether verbal or propositional. To the extent that it is, the definitions are implicit, and I do not believe I differ from others in failing to observe any such definitions introspectively. If some claim otherwise, I would challenge them to distinguish the observation of such a definition from the observation that one is simply disposed to assent to the (verbal) definition. When, later in this book, I will defend the claim that an equivalence of the kind mentioned in Thesis (I) exhausts speakers' semantic interpretation of "true", I do so for the reason that no more seems needed to explain the data, and that the simplest theory is the best. Thus, speakers do not need to be able to detect this principle as giving the meaning to "true"; this seems not to be an appropriate requirement on any semantic claim.

Returning now, finally, to the issue of descriptivity, the upshot of the above line of thought is that because "true" is most plausibly taken to be

*un*ambiguous, a common reason for engaging in technical definitions – ambiguity – cannot be appealed to here. This gives us good reason to say that there is a unique concept of truth, and that any definition of a predicate which makes it differ semantically from our ordinary truth-predicate would fail to capture that concept, and so fail as a theory of truth.

The second way in which the analogy with "red" may be disputed is by saying that "true", like "fish", "set", or "electron" can be redefined with theoretically beneficial consequences. If so, there would be a different reason to speak of a manifold of truth-concepts and to make use of technical interpretations of the word. But, first, "fish" is a bad example here, for it was surely misleading, rather than enlightening, to redefine "fish", instead of inventing a new word for the intended family of maritime creatures, and then conclude that fish can have very different ancestries. It is of course common to hear that "science has shown what fish really are". This is just an authoritative *use* of the very word under discussion, and not a cogent argument. And one cannot appeal to externalist theories of natural kind terms, for "fish" never denoted a natural kind, but a so-called "mixed" kind, like "sand" or "jade". The scientific concept of fish is of course better, relative to many important *desiderata*, but this also has little relevance for the terminological question. Next, "set" is a word which is typically introduced merely to define further notions, or to enable certain expressive strengthenings, and is not taken to answer to any intuitions that seem philosophically important. It is doubtful whether "electron" was ever defined or redefined at all, rather than merely introduced with a number of synthetic hypotheses.

In any case, the differences with "true" are rather striking. When philosophers throughout history have pondered truth, seen connections between truth and other concepts, like knowledge and evidence, and asked "What is truth?", surely, they were wondering about the ordinary notion of truth and not some refined, distinct notion. Further, "true" is a prescientific word of natural language, and has not, like"electron", been introduced as a posit in order to explain certain phenomena. It does not make sense to say that it should be redefined in order to "better accommodate the data", for there are no data which the term has been introduced to explain. This is not to deny that a concept sharing many (but not all) properties of truth, as ordinarily conceived, e.g., the predicate Tarski defined, may be useful for some purposes. (Note that this is something Davidsonians must deny – they cannot consistently take Tarski's definition to be a good theory of truth, only a good theory of meaning.)

As with "fish", it is only misleading to say that the stipulative definitions are definitions of truth. In which sense of "truth" is it a definition thereof? It is clearly false to say "In the ordinary sense", and trivial to say "In the sense of the definition". To say that the ordinary notion of truth is "defective" and that we therefore "must" define a new one is confused. If the ordinary notion of truth is defective, a correct theory about it should describe it as defective. If for a certain theoretical purpose, one needs a predicate with some but not all the properties of the ordinary truth-predicate (presumably one lacking its defective features), then one is free to define one. But it is wrong to say that the motivation for this project is that "we need the concept of truth" for such and such. Clearly, it is not truth one needs.

To this, it might be replied that truth is a property which our ordinary truth-predicate *purports* (but fails) to pick out. If we can say, then, what this property is, and where our ordinary "true" goes wrong, there is reason to deny the constraint of descriptivity. But if one takes "true" to *fail* to pick out the property in question, one may not, for the obvious reasons just rehearsed, use "truth" and take it to be understood as it is ordinarily understood when speaking of this property that "true" purports, but fails, to express. Finally, it may be thought that there are some principles governing the meaning of "true" which determine it as expressing the property of truth, while there are others that do not, and which make the word "defective". Here, it is cognitive, or semantic, meaning that is relevant, not connotation, poetic value, etc. But it is difficult to see how principles determining the semantic meaning of a predicate can fail to affect which property it expresses. Surely, such principles determine the conditions for something to have the property the word expresses. But then, it also determines which property it expresses.

So, none of these two ideas are of any help in legitimizing the claim that "we need truth for such and such, but we need to refine the notion". A refined notion will *ipso facto* not be truth, and, therefore, the idea that truth is needed for some purpose can only be coherently interpreted as the claim that truth, with all its alleged defects, is needed. The plausible retreat position is that some notion similar to truth is needed. Of course, all arguments for the need of a refined notion are arguments which exploit our ordinary understanding of "true", and, so, only show, if anything, that the ordinary notion is needed. One wonders what could possibly count as an argument for the retreat position. So, the common claim that "if the ordinary notion of truth is defective, we must define a new one" turns out to depend on a ulterior premise about what aim we have. If we want to study the ordinary notion, the

claim is obviously false, and if we are interested in applying a redefined predicate for some reason, then we must not appeal to arguments that truth is needed. Ironically, it is common to take Tarski style, recursive truth-definitions to be good regimentations of "true", while taking the important purpose of "true" to be in semantics and logic, although, as Dummett made clear long ago (1978: xxi), Tarskian truth-definitions cannot both be definitions of truth and semantic theories of a language.

Let us now look at the three further constraints, belonging to the obvious constraints that any theory must meet, such as consistency and lack of ambiguity. I will call the first of these the *constraint of unification*. We have already noticed some facts about the truth-predicate: that sentences of the form (i)-(iii) are equivalent to the substituting sentence "*p*", that (1) implies all instances of the form "If he said that *p*, then *p*", that "What he said is true" implies all those of the form "If what he said that *p*, then *p*" and that arguments like (B) are valid. There are also the "facts about truth", though there may be some disagreement as to what is and what is not a fact. Now, a theory of truth cannot just state these facts about "true" or about truth proper as separate rules, because it would then not be possible to see what it is about the truth-predicate that had all this as consequences. Thus, the theory must be *unified*, meaning that all diverse facts about truth and occurrences of "true" be explained by recourse to the same (non-conjunctive) statement. It would not do, for instance, to take the conjunction of (ES), (PS) and (DS) to be the theory of truth. This is quite obvious and a kind of simplicity-constraint that all theories must meet. Still, many of the theories that have been called deflationary do not respect it at all, and their originators seem not to have thought of it when developing their theories. It therefore seems appropriate to state the constraint explicitly, in all its obviousness, and for quick and convenient reference later on.

The two last general constraints that are especially relevant to deflationism are those of *non-circularity* and *finite formulation*. These are considered here because it has been suspected that deflationists cannot meet both these constraints simultaneously. This is for the reason that we do not want merely to claim that snow is white iff snow is white, that it is true that grass is green iff grass is green, and so on, and say that this is what explains truth. We want to generalise over these particular facts, viz., to give a finite claim that implies the particular instances. But it has been difficult to see how this could be done in any other way than by saying that all instances of (ES) are *true*. But such a solution, of course, raises the suspicion of circularity.

We must first ask whether we really must accept the constraint of finite formulation, since some deflationists, notably Horwich, have accepted an infinitely formulated theory. Horwich's theory, which he calls the Minimal Theory (MT), contains as axioms all instances of the schema (PS) and these axioms are not derived from any other finite collection of claims. Horwich admits that the infinitary character of his theory is a weakness, but he considers it inevitable (1998: 25ff.). The reason such infinitary solutions are unacceptable, I will argue, is that it would be unreasonable to believe in deflationism in the first place if it turns out that no finite claim covers the equivalences, and which, by implication, explains everything about truth. It is simply a general criterion of all theories that they find a law explaining the instances among the *explananda* that share a certain feature. The common feature of the facts that the deflationist appeals to is the form of the sentences expressing them. If no deflationary theory can derive these by recourse to some law, we should believe that some other theory can, and then we should try to find that theory, and abandon deflationism. This is not to say that there are no infinite theories that are very fruitful and explanatory. Obviously there are, e.g., in mathematics. But we are now concerned with the philosophical question about truth, *a fortiori*, whether a certain kind of theory can explain *everything* about it. Therefore, we should take the demand for finitude differently from in other cases.

Theories are for explaining things. That is why we should expect something finite and law-like, rather than an infinite theory from a final, exhaustive theory of truth. Not only the explanation of facts about truth becomes implausible on an infinite theory, but the claims of the theory itself seem to be in need of explanation. To wit, on a theory about a certain property, it cannot simply be taken as primitive that, e.g., a certain proposition has this property depending on the colour of snow. We would like to know *how* this can be. It also seems implausible that the theory should use all possible concepts, including that of phlogiston and *Dasein*. There is nothing wrong with using concepts with empty extensions, of course, since they may be formed out of perfectly legitimate ones (like square and round), but one should not be forced to endorse any concept as if they were all legitimate (cf. Gupta (1993a: 365)).

Infinite theories are more vulnerable in their ambition to explain other facts about truth. The defender of an infinite theory will have to claim that what explains why, e.g., true beliefs facilitate successful behaviour is (partly) that the proposition that snow is white iff snow is white, the proposition that

grass is green is true iff grass is green, and so on *ad infinitum*. But, first, the "and so on" can be interpreted in infinitely many unintended ways. It must therefore be taken as somehow standing proxy for the whole, infinite continuation of the series, which is an infinite conjunction. But, secondly, even if, *per impossibile*, we could give the infinite claim in giving the explanation, that would not be acceptable anyway. When we want to explain something, we want to do it by a finite number of other general claims, not an infinite number of particular claims. That seems to be in the very nature of an explanation, and accordingly, all philosophical attempts to analyse this notion refers essentially to laws or generalizations.[14]

It may seem, however, that (MT) does not need to fare this badly, but that one can easily formulate the matter so that the *explanans* is not infinitary in the above sense. Could one not say that the perfectly generalizing and finite claim is that all the instances of (PS) are axioms of the Minimal Theory? No, for we cannot from this claim infer the instances themselves, or use them to explain any facts, unless we also assume that the axioms of the Minimal Theory have some specific property (e.g., that they are all true or have some other feature allowing us to use them as premises). It is obvious that if nothing more is said about (MT), then such an explanation is inadequate, since the fact that the instances are axioms of (MT) cannot explain anything about truth, since this fact is simply made true by definition and therefore empty. Rather, Horwich's claim must be taken to be that the Minimal Theory itself explains truth. But then, we arrive again at the claim that, e.g., the fact that true beliefs facilitate successful behaviour, is to be explained (in part) by this: that the proposition that snow is white is true iff snow is white, and so on. But this is precisely the unacceptable consequence we noted in the first place.

It is important not to conflate the constraint of finite formulation with that of unification. One can obviously violate the latter without violating the former. Since, for instance, Horwich's theory follows a very simple pattern, in all its infinity, the converse also holds. Thus, if Horwich manages to derive all and only the theses he wants, then he will have met the constraint of unification, while violating that of finite formulation.

---

[14] Another well-known explanatory weakness of (MT) is the difficulty of explaining general facts, rather than merely the instances thereof (cf. Gupta (1993b: 66) and Soames (1999: 247f.)). In 2.12, we will discuss this problem as well as Horwich's proposed solution to it, and in 5.5, I will present my own solution to this problem.

One may perhaps also question the constraint of non-circularity, or the claim that saying that such and such instances are true would violate such a constraint. This will be the topic of 3.2. The upshot of Chapter 3 as a whole is a solution to this problem of meeting both the constraint of non-circularity and that of finite formulation, the problem that I shall call the *problem of formulation*.

CHAPTER TWO:
A CRITICAL HISTORY OF DEFLATIONISM

2.1 INTRODUCTION

Though the schema-instances appealed to by deflationists were discussed already in antiquity (e.g., by Aristotle in *Categories* 14^{b}15), it is not until fairly recently that anything resembling a deflationary theory has been vented. The aim of this chapter is to give a fairly extensive presentation of those ideas that, for one reason or another, can be thus characterised. The views of philosophers with some connection to deflationism, from Frege to Horwich, are presented and commented upon, with old objections rehearsed and new ones formulated. Since in Ch. 3 and onwards, my own ideas will be developed, it must first be said why (most of) the alternatives must be rejected. The chapter is also supposed to serve as a person-oriented introduction to the subject of deflationism, as opposed to the static account of its logical space found in Chapter 1.

2.2 FREGE

Frege, though not unambiguously a deflationist, can be regarded as the first philosopher to express deflationary views about truth. However, the scarce comments on the matter are not entirely decisive and seem also to be in some tension. It is reasonable to regard them as pieces of speculations, rather than arguments intended to support a final view.

One *locus classicus* of his ideas about truth is "The Thought", where Frege writes:

> It may […] be thought that we cannot recognize a property of a thing without at the same time realizing the thought that this thing has the property to be true. So with every property of a thing is joined a property of a thought, namely, that of truth. It is also worthy of notice that the sentence 'I smell the scent of violets' has just the same content as 'it is true that I smell the scent of violets'. So it seems, then, that nothing is added to the thought by my ascribing to it the property of truth. And yet is it not a great result when a scientist after much hesitation and laborious research can finally say 'My conjecture is true'? The meaning of the word 'true' seems to be altogether *sui generis*. (1918: 36f.)

Another is in "On Sense and Reference":

> One can, indeed, say: 'The thought, that 5 is a prime number, is true'. But closer examination shows that nothing more has been said than in the simple sentence '5 is a prime number'. (1892: 64)

(Frege uses "thought" instead of the now more common "proposition".) None of these passages express explicitly either (I) or (II), but since Frege takes the strongest possible equivalence to hold between the truth-ascriptions and the denominalized sentences, he would perhaps not be wholly foreign to such views, i.e., that the equivalence explains the concept of truth.

As Frege himself notes in "The Thought", though it is easy to explain the truth-operator, just by saying that it confers no new content to sentences prefixed by it, it is uncertain how to explain other occurrences of "true" in a way consistent with this redundancy. In particular, Frege's hesitation concerning the sentence "My conjecture is true" arises precisely because no sentence lacking "true" can be seen as trivially equivalent to it, as was the case with the truth-operator. This interpretation is due to Scott Soames (1999: 21ff.) and seems to be the most plausible explanation of Frege's concerns about the alleged "great result" of the scientist. This is a problem which, as we will see, has been given different kinds of solution. But whatever we say about it, it is clear that Frege touched upon a central issue of deflationism: the difference and relation between redundant and irredundant occurrences of "true". However, Frege did not himself provide any solution to the problem.

What, now, of the claim that one "cannot recognize a property of a thing without at the same time realizing the thought that this thing has the property

to be true"? This is reminiscent of claims that have been made more recently about the connection between truth and assertion. It has been held, e.g., that one cannot assert that a thing has a certain property without claiming the truth of one's assertion. Not everyone has seen this connection between assertion and truth as a motivation for deflationism – on the contrary (Wright (1992: I.III), Cozzo (1994: 155f.) and possibly Dummett (1973: 459f.)). For Frege, however, this idea is really only a trivial consequence of his other claim, that sentences of the form "The thought/proposition that p is true" express the same thought as the corresponding "p", together with the plausible claim (which Frege held) that synonymous expressions are interchangeable with no change of truth-value. For the sentence "X recognizes that (the thing) a has (the property) b" will then express the same thought as the corresponding sentence "X recognizes that the thought that (the thing) a has (the property) b is true". The opening claim of the passage from "The Thought" is therefore redundant.

What, now, of Frege's other comments on truth, and their relationship with the two deflationist passages just quoted? Frege comes, at various places, to three conclusions which are among the central theses of most deflationary accounts of truth, namely: (1) that truth is not a property (1892: 64); (2) that truth is indefinable (1918: 35); and (3) that the correspondence theory of truth is unworkable (1918: 34f.). These are all negative claims, but claims which, as we have seen, are intimately connected to the deflationary theory of truth. We will now look a little closer at the relationships between (1)-(3).

The first claim Frege tries to support by a well-known argument which has been quite decisively refuted. It holds that since the primary fact about truth is that there are two *objects*, The True and The False, which are the referents of thoughts, to speak of a *property* of truth is misleading. The argument goes:

> One might be tempted to regard the relation of the thought to the True not as that of sense to reference, but rather as that of subject to predicate. One can, indeed, say: 'The thought, that 5 is a prime number, is true'. But closer examination shows that nothing more has been said than in the simple sentence '5 is a prime number'. The truth claim arises in each case from the form of the declarative sentence, and when the latter lacks its usual force, e.g., in the mouth of an actor upon the stage, even the sentence, 'The thought that 5 is a prime number is true' contains only a thought, and indeed the same

thought as the simple '5 is a prime number'. It follows that the relation of the thought to the True may not be compared with that of subject to predicate. Subject and predicate (understood in the logical sense) are indeed elements of thought; they stand on the same level for knowledge. By combining subject and predicate, one reaches only a thought, never passes from sense to reference, never from a thought to its truth value. One moves at the same level but never advances from one level to the next. A truth value cannot be a part of a thought, any more than, say, the Sun can, for it is not a sense but an object. (1892: 64)

Frege thought that he had certain technical reasons for postulating these objects, but not many today finds the notions of The True and The False credible. The common view is also that Frege was not even forced by his own theoretical commitments to view truth in such a way, at least not by those commitments worth hanging on to. Furthermore, even if we would grant that the relations between The True and thoughts are the fundamental facts about truth, we could still define the property of truth as the property that a thought has when and only when it refers to The True.[1] It is thereby also clear that this piece of reasoning does not (as opposed to Frege's intentions) lend any additional credibility to a deflationary view of truth.

The second claim is somewhat ambiguous. It may mean either that there is no way at all to define truth, or it may mean that one can give the meaning of "true" by citing the equivalences between "The thought that p is true" and corresponding "p", etc., but that it is *not* possible to give necessary and sufficient conditions for something's being true, viz., to give a truth-analysis of the (TA)-form, hence, an endorsement of thesis (III). If this latter interpretation is correct, it supports the view that Frege was a deflationist, for the denial of the possibility of an analysis of the form of (TA) is a typical motivation for deflationary theories. If the former interpretation is correct, however, it would be wrong to call Frege a deflationist, for a deflationist typically presents a claim precisely about how to give the meaning of the truth-predicate. In order to judge which of these two views he held, we must discuss the third claim, that the correspondence theory is unworkable. One argument for this claim goes:

We should have to inquire whether it is *true* that an idea and a reality, say, correspond in the specified respect. And then we should be confronted by a

---

[1] Cf. Blackburn (1984: 228f.), Soames (1999: 47) and Pagin (2001: 3).

question of the same kind, and the game could begin again. So the attempted definition of truth as correspondence breaks down. For in a definition certain characteristics would have to be specified. And in application to any particular case the question would always arise whether it were *true* that the characteristics were present. So we should be going round in a circle. It therefore seems likely that the content of the word 'true' is *sui generis* and indefinable. (1918: 35)

Dummett (1973: 442ff.) has argued quite conclusively against the soundness of this argument. In particular, Dummett argues that the argument is unsound unless in order to determine the truth of any sentence *s*, one must first determine that of the sentence saying that *s* is true, which seems incorrect (see also Soames (1999: 24ff.) for a detailed treatment of the argument). What is odd, further, about this argument is that, though directed against the correspondence theory in particular, it would, if sound, prove that no explanation of truth is possible whatsoever. For in order to "give the meaning of", or "define", the word "true" by claiming, e.g., (B) above, then we would have to claim that (B) is true, which would, according to the argument, launch an infinite regress. But if he accepted such a conclusion, he would be holding (2) on the stronger interpretation, which would count against the interpretation of Frege as a deflationist.

In recapitulation, we may say that there seems neither to be any conclusive evidence showing whether Frege was a Deflationist or not, nor does there seem to be a simple way of explaining precisely in what sense Frege thought truth to be *sui generis*. It appears reasonable, rather, to regard this claim of uniqueness as result of the variety of conflicting (and often confused) views that he held about the matter. However, it is clear that some of his claims were unambiguously deflationist, while others go beyond or even against deflationism.

2.3 THE REDUNDANCY THEORY – RAMSEY AND WILLIAMS

Frank Ramsey is usually thought to have given the first unambiguously deflationary account of truth[2], one which is commonly called the Redundancy Theory, in his article "Facts and Propositions" from 1927. As the name of the theory suggests, Ramsey argues that "…is true" adds nothing to what it has been applied to. "It is true that Caesar was murdered" says nothing more than "Caesar was murdered" (1927: 157-8). The same holds, according to Ramsey, when, instead of using a truth-operator, we apply the truth-predicate to a name of a sentence. Thus "'Caesar was murdered' is true" says only what "Caesar was murdered" says (1927: 158).[3]

The redundancy claim here seems to be that for any sentence containing the word "true", one can find a paraphrase lacking it. Thus, Ramsey claimed that the truth-predicate was redundant not only in such simple sentences as the left-hand-sides of (ES) and (DS), but everywhere. Recall that Frege was troubled precisely about those sentences containing "true", which did not contain an embedded sentence that could itself be taken as the appropriate paraphrase, such as "My conjecture is true". Such sentences are called "blind ascriptions of truth". Another example is the quantified sentence (1) above. As seen above, these sentences are interesting in that it is thanks to them that "true" increases the expressive power of our language. One would like to say that it is in such sentences that "true" is not redundant, but in a special sense, it is, namely if you allow the paraphrases to contain higher-order quantifiers. For the quantified sentence (1), for instance, Ramsey suggested the following paraphrase:

(1a)   For all *a*, *R*, *b*, if he asserts *aRb*, then *aRb*.

However, since there is no reason to give sentences of the form "*aRb*" any primary status, it has been common to use instead propositional quantifiers,

---

[2] This is doubtful, however, for both Frege and W.E. Johnson (see his (1921: IV, §2)) are plausible candidates for being the inventor of deflationism. The biconditional itself was, to my knowledge, noted first by Aristotle (*Categories* 14[b]14-18), but he did not claim that it be taken as definitional or exhaustive of truth.
[3] Ramsey's very short passage does not use quotation marks, but I here include them for the sake of grammaticality – cf. note 6 of Chapter 1.

that is, quantifiers which bind into sentence position. Thus, instead of (1a), it has been thought that the paraphrase of (1) should be:

(1b) (p)(If he said that p, then p).

Since there is no *prima facie* reason to regard quantification into sentence position as more objectionable than quantification into predicate-position, and since the gain of (1b) over (1a) in generality is obvious, I shall concentrate on the propositional variant. Ramsey did not say more on behalf of this redundancy theory than this, but many have discussed its implications and merits, but I will restrict my discussion to the more elaborated ideas of the second redundancy theorist, C. J. F. Williams.

In contrast to Ramsey's laconic comments on truth, Williams has devoted a small book (1976) to the subject. Its main theme is to find paraphrases of such blind ascriptions as Frege's example "My conjecture is true", though Williams' favoured example is the sentence

(2)   What Percy says is true.

The method here is to apply a kind of variant of Russell's analysis of definite descriptions, and the end result (1976: 38) is:

(2a) $(\exists p)(q)((p = q \equiv$ Percy says that $q)$ and $p)$.

Now, if we take the Redundancy Theory to be saying that every occurrence of "true" can be eliminated by paraphrase, then it should come as no surprise that its proponents have made so much use of propositional quantification. The reason is that there seems to be no way to achieve the paraphrasing, while still engaging in what may properly be called a deflationary theory, without it. (This last condition is mentioned for the obvious reason that if we are not deflationists we can easily eliminate an occurrence of true, simply by replacing it with "corresponds to the world", "is satisfied by all objects", etc.) Given a sentence like (1), the rendering "(p)(If he said that p, then p), or something similar, seems to be the only candidate for a paraphrase.

Evidently, then, much of the project of claiming the truth-predicate "redundant" in the sense of "eliminable" hinges on the possibility of using propositional quantification. But our natural language does not contain any such kind of quantifier (as will be argued at greater length in 3.3). This is the

point mentioned above: without the truth-predicate, we would have to invest in additional grammatical constructions, which, for some reason, our language tends to avoid.

The quirk now comes to this: if we did have something like propositional quantification in our language, the truth-predicate would be redundant in three ways: first, it would not be needed for forming the blind ascriptions (thanks to which "true" increases the expressive power of English) – this would be achieved by the quantification; second, it would be eliminable by paraphrases, thanks to the propositional quantification. But these two points surely come to one. This shows once again why "true" increases the expressive power only of a language which lacks propositional quantification. The third reason why "true" would be redundant given that we may use propositional quantification is that we could then give an eliminative definition of "true", i.e., a truth-analysis of the (TA)-form. One example of such a definition, taking propositions as primary truth-bearers, could be this:

(TA3) For all x, x is true iff $(\exists p)$ x is the proposition that p and p.[4]

We will discuss at length the problems of using propositional quantifiers in framing a truth-theory in 3.3, but we may note here that, just as with (TA1), this proposal as a theory of truth seems only to introduce a new notion, the explanation of which would provide us with the means for giving a more direct theory of truth, i.e., one which does not require for its completion any definition of further expressions.

Now, the Redundancy Theory is correct in this way: the paraphrases, if the propositional quantifiers are interpreted in the intended way, are indeed equivalent to the original sentences in the sense that the original sentences and their paraphrases have the same inferential properties. The problem with the Redundancy Theory, however, is that it seems unnecessary to give the paraphrases instead of just saying what different types of sentences are implied by different kinds of truth-claims. The paraphrases do not tell us anything more of interest than this. More importantly, the redundancy theorists can be accused of not having given a unitary explanation of "true" (cf. Kirkham (1992: 324f.)). In particular, how do the paraphrases relate to the claim about the equivalences between the halves of the instances of (ES) and (DS)? Further, the paraphrases do not show what sentences like (1) and

---

[4] Cf. David (1994: 74ff.), Baldwin (1989) and van Inwagen (2002).

those like (2) have in common. There is no fact or rule appealed to which determines the various paraphrases as the correct ones. The Redundancy Theory therefore fails to meet the constraint of unification

Further, we said that a truth-theory is not required to give paraphrases lacking "true" of sentences containing it. But paraphrases may also be misleading, if taken to say more than they really do. Williams clearly regarded them as not only semantically equivalent to the paraphrased sentences, but also to display the real *logical form* of the truth-claims. This goes counter to the plausible idea that natural language has invested in new expressions of old grammatical type, but with new semantic properties, in order to achieve the increased expressive power in a "cheap" way. But it is also a directly implausible idea that quantified truth-claims should take some other form than other quantified sentences, viz., that "Everything he said is true" should be of different form than "Everything he owns is expensive". In any case, between two rival theories, we should *ceteris paribus* choose the one that does not multiply the interpretations of the logical form of superficially similar sentences (these issues will be more carefully treated in Chapter 5).

From this idea about the logical form of truth-claims, Williams has drawn some serious philosophical consequences. One such consequence is that truth is not a property and there is no such thing as a truth-bearer (1976: Chs. 3 and 4). The argument for the first claim is that "is true" is not a predicate, since the "correct" analysis of sentences containing it is such that no truth-predicate occurs in the *analysans*, viz., the propositionally quantified paraphrase (1976: 17ff.). A corresponding argument is put forward for the second claim. But, again, how can it follow that "is true" is not a predicate from the fact that, given an explication of propositional quantifiers, a certain sentence lacking the truth-predicate can be shown equivalent to a sentence containing it? Some further premise about logical form is needed here, but one we have good reasons to resist (cf. Kirkham (1992: 71f.)).

In conclusion, Ramsey's paraphrases were very instrumental in leading to the deflationist insight into the semantic functioning of "true", although at least Williams, it has been argued, went too far in treating them as something more like descriptive analyses of the various paraphrased sentences. Most importantly, the redundancy theorists have not presented a unified theory about "true", and must therefore be regarded as incorrect.

2.4 TARSKI

Philosophers seem rather divided as to whether Tarski's type of truth-theories should be regarded as a deflationary or not. Though this is to some extent a terminological issue, I will argue that on the most appropriate use of "deflationism", Tarski's standard theory is non-deflationary, while there are non-standard versions that are more deflationary in spirit. I will discuss those alternative versions below, and focus first on the standard interpretations. At the end, the standard theory will be criticised, partly in the same the way that correspondence theories were criticised in 1.3.

A distinctive feature of Tarski's view of truth is his insistence upon a certain adequacy condition for a truth-theory, which he calls the *criterion of material adequacy*. A truth-definition meeting this criterion is supposed to be guaranteed to capture the extension of the truth-predicate (1944: 353f.). It holds that a truth-definition must entail every sentence of the form "*S* is true iff *p*" (a "T-sentence"), where "*S*" is a name or description of a sentence and the sentence replacing "*p*" is that very sentence or a sentence with the same meaning. The former will hold when the language that the T-sentence belongs to (the "metalanguage") is a superset of the language to which the sentence named in the left hand side of the T-sentence belongs (the "object language"). The latter will hold when the two languages are disjoint (1944: 344, 350). Tarski believed that any language that contains a predicate that applies to all and only the true sentences of the language is inconsistent. This is because of Liar-sentences like "This sentence is false", which can be shown to be both true and not true. He therefore forbade this to happen, by imposing the restriction that the object language always be at most a proper subset of the metalanguage. Specifically, a truth-predicate for the sentences of the object language cannot be in the object language, but must be in the metalanguage (1944: 348ff.).

The standard Tarskian truth-definitions operate with the notions of *satisfaction* and *denotation*. Satisfaction is a relation between objects and *open sentences*, such as "*x* is wise", i.e., sentences with free variables. Though satisfaction is not explicitly defined, it is to be at least preliminarily characterized by examples of what satisfies what. For instance, Socrates satisfies "*x* is wise" since Socrates is wise, Fido satisfies "*x* is a dog" since Fido is a dog, and so on. Thus, the notion of satisfaction is supposed to be explained so as to make true all instances of the schema "*a* satisfies '*F(x)*' iff *F(a)*". Similarly with denotation, which is explained by the claim that, e.g.,

"Aristotle" denotes Aristotle, and so on, i.e., it is to validate the schema " '*a*' denotes *a*".

Tarski then defines truth for sentences of a first-order language *by recursion over the syntactic form of the sentence*. On a very simple version of such a definition, this means that the *definiens* is a long disjunction such that for each form a sentence can have, there will be a disjunct ("clause") saying under which conditions a sentence of that form is true. For instance, one disjunct will be "*s* is of the form '*p* or *q*' and is true iff one of its disjuncts is true". Thus, if a given sentence is complex, i.e., is formed by two sentences with a connective combining them, its truth-conditions will be given by the clause for that connective in terms of the truth-condition of its constituent sentences. If these are atomic, then their truth-conditions are given without recourse to truth. If not, then their truth-conditions are again given by the clause corresponding to the form of the sentence. This step is repeated until one reaches the embedded atomic sentences, whose truth-conditions determine those of the complex sentences containing them, which in turn determine those of the sentences containing them, and ultimately, the given sentence itself, whatever its complexity. The truth-conditions of the atomic sentences are also given by clauses covering their form. For instance, one such disjunct will deal with universal sentences, which says that a universal sentence "$(x)...x...$" is true iff *every* object satisfies the open sentence "$...x...$". Further, an existential sentence "$\exists x...x...$" is defined as true iff *some* object satisfies "$...x...$". When an open sentence here is complex, the conditions under which an object satisfies it may again be given recursively, by clauses such as "*o* satisfies '$F(x)$ or $G(x)$' iff *o* satisfies '$F(x)$' or '$G(x)$' ". An atomic sentence "$F(a)$", finally, is true iff the object that "*a*" denotes satisfies the open sentence "$F(x)$". I should hasten to say that this is a very primitive and in some respects defective type of truth-definition, but it illustrates the general idea sufficiently for present purposes.

Now, was Tarski a deflationist? The main reason to answer affirmatively is that if (1) the "names or descriptions of sentences" referred to in the criterion of material adequacy are quote-names (which usually is the case), and (2) the metalanguage = the object language + semantic vocabulary, then the T-sentences will be instances of (DS) (see Thesis (II)). Indeed, the recursive definitions have been seen as one way of solving the problem of formulation discussed in 1.5, the problem of generalising over the instances. One may, however, regard the way Tarski does this as non-deflationary, claiming that the generalisation must be more "direct", e.g., as in thesis (I).

This could be motivated by saying that if the derivation is so indirect as in the recursive definition, then the theory does not capture the idea that for "Snow is white" to be true is just for snow to be white, etc. One can explain this intuitive notion of directness by saying that a claim C is a direct generalisation over the sentences of a set S to the extent that it entails all and only sentences in S with *few* conceptual resources. Tarski's use of satisfaction, etc., thereby makes it clearly less direct than the deflationary theories we have been considering. Further, it seems to entail significantly more than the equivalences. Many Tarskian definitions, for instance, entail that various names denote certain objects and that various open sentences are satisfied by such and such objects. Tarski rejected a direct theory like the Redundancy Theory because he required that each occurrence of "is true" be eliminable, and therefore sought for an explicit definition, i.e., one giving necessary and sufficient conditions for a variable truth-bearer to be true. This, of course, makes it non-deflationary on the interpretation of "deflationism" as containing thesis (III). Another common reason for taking Tarski to be a deflationist, rather than a correspondence theorist, is that his theory does not operate with the notion of fact or state of affairs. However, this cannot be sufficient for being a deflationist, since epistemic theories would then be deflationary. Finally, the most important reason why Tarski should not be regarded as a deflationist is that he makes use of the notions of satisfaction and denotation, which are representational notions, i.e., notions of relations between truth-bearers and objects in the world.

The question whether Tarski is a deflationist may now seem utterly terminological, but I will argue that there are good reasons for adopting a terminology according to which Tarski's standard theory is non-deflationary. As we have seen, the only good reason not to is his insistence on the criterion of material adequacy. But it cannot be right to say that any theory self-consciously meeting this criterion is deflationary, because if the instances of (DS) are true, then this must be explained by *any* theory of truth, including correspondence theories. If the T-sentences are not true, then any theory meeting Tarski's criterion is false, so that possibility need not be considered. Thus, it seems appropriate to use a terminology according to which a theory meeting the criterion of material adequacy may, but need not, be deflationary. Considering this, and Tarski's use of representational notions and the indirect nature of the derivation of the T-sentences, it seems reasonable to say that Tarski's standard theory is non-deflationary. Though it may be argued that Tarski, at some point, meant that the T-sentences are exhaustive of truth, this

is reasonably seen as contradicted by the subsequent endorsement of definitions that entail these in the indirect way noted above.

As mentioned above, there are versions of Tarski's theory that are more deflationary in spirit. Since Tarski considers as adequate any definition which fulfils the criterion of material adequacy for a given language, one can, for the very simple language containing only the sentences "It is snowing" and "It is raining", give as a correct truth-definition for that language "s is true iff (s = "It is snowing" and it is snowing) or (s = "It is raining" and it is raining)". Given the criteria that the theory not use representational notions and that the derivation of the instances of (DS) be direct, this theory perhaps qualifies as deflationary. But, of course, such definitions are uninteresting, given their narrow scope. One could, however, generalise the idea using a propositional quantifier, as in "For all s, s is true iff $\exists p(s = 'p'$ and $p)$", which was considered but rejected by Tarski. This is a type of deflationary theory that has often been considered, and which will be discussed in 3.3.

Now for the critical assessment of the standard Tarskian theory. One weak point here concerns the use of the notions of satisfaction and denotation. Clearly, the truth-definitions are no less problematic than these notions. In general, I think the problem about Tarski's use of satisfaction is better appreciated when it is realised that one may substitute for "satisfies" the dummy-predicate "φ:s", together with the claim that Socrates φ:s "x is wise", since Socrates is wise, and so on, or a list saying what particular objects φ:s what open sentences, and you have an equally informative account of truth. (Indeed, I think a fair assessment of Tarski's theory is that, as long as satisfaction is not explained, the Tarskian definitions should be regarded rather as definitions of the notion of satisfaction in terms of truth, rather than the other way around. This is because "true" is understood, whereas "satisfies" is an undefined technical term, which may very well turn out to be either inexplicable or inadmissible to use to define truth.)

There are now three ways of understanding the way Tarski uses these notions, and, as will be argued below, problems emerge on all of them. First, one can say explicitly between which objects these relations hold in a list, and give a truth-theory for only such languages for which such a list is provided. (This is what Tarski actually does in his definitions.) Second, one can take them to be fully explained by examples as above. Third, one may leave it for a future theory to explicitly define the notions for variable expressions and objects. As will now be detailed, the first two options seem to make the truth-theory seriously incomplete (if in quite different respects),

while the third will make for a quite ambitious programme with considerable difficulties. Evidently, the truth-theory will inherit all difficulties facing that programme. There is also a more philosophical worry that even if the programme can be executed, the theory will not have explained what it is to be true.

The unattractive feature of the first treatment of the problematic terms is that the truth-definition will be restricted to specific languages. As argued against the epistemic truth-theory of mathematical truth, a fully general truth-theory on these lines seem to make the notion of truth disjunctive in an unacceptable way. But this theory would make "true" not only ambiguous, but indefinitely so. For any conceivable language, there must be a clause pertaining to the truth-theory which specifies the truth-conditions for every sentence of that language. On this option, further, for any new language, one could not extract from the definition what it is for a sentence in the language to be true, but would need an addition pertaining specifically to that language.[5] One would also have to regard as distinct languages "English" as spoken before the introduction of a given new expression and English after the introduction (cf. Black (1948: 57f.)). A descriptively correct (i.e., not stipulative) truth-theory, by contrast, should say what true things have in common, in virtue of which they are true. Without a general account of the representational notions, this cannot be achieved with Tarskian definitions. (As will be seen, however, Soames (1999: 110ff.) has shown that accounting generally for the notions of denotation and satisfaction will not solve the problem, since each type of logical constant (conjunction, universal quantifier, etc.) must also be independently characterized, viz., not in terms of truth.) It should be noted that the difficulty of giving a truth-theory for sentences of variable languages seems to afflict many truth-theories which take sentences as primary truth-bearers. It seems to me that the only way of avoiding it is to appeal to some translinguistic notion of fact, state of affairs, proposition, belief, etc.

On the second treatment, the paradigmatic examples of things related by the two relations give a sufficient explanation. There are well-known objections to this type of treatment, the most important of which is probably that the examples do not by themselves determine what further things are or are

---

[5] This lack of generality, the fact that "true" is not defined for a variable language, is probably the most common objection to taking Tarskian definitions as theories of what it is for a sentence to be true (cf. Black (1948: 56ff.)).

not related by the relations. The series can be continued in indefinitely many unintended ways, in the way the later Wittgenstein famously illustrated (1953: I: §§ 185-87). This does not show that the notions cannot be given the desired general explanation, of course, but only that some such explanation must be possible if the theory is not to use notions that may seem unfit for taking as primitive.

This conclusion is perhaps more worrying for Tarskian truth-theories than it first appears. It may appear that we have a firm grasp of the notions, since we can easily tell what the conditions are for a given object to satisfy an open sentence. The problem is that it may turn out that when it is realised how we understand the notions, we see that a corresponding explanation thereof is incompatible with the truth-theory being adequate. To see this, recall first that the notion of satisfaction is supposed to validate the schema "*a* satisfies '*F(x)*' iff *F(a)*" and that denotation is to validate " '*a*' denotes *a*". There are now two ways in which our understanding of these notions could defeat the theory. Either, it turns out that we understand the notions in terms of truth, so that the theory is circular. This would be so if our understanding of the notions is essentially connected to our taking the instances of the schemata to be true. (Circularity would also result if our understanding of satisfaction is given by an analysis like: *o* satisfies "*F(x)*" iff for some term *t*, which denotes *o*, "*F(t)*" is true (cf. Tarski (1944: 353)). The fact that "true of" is sometimes used synonymously with "satisfied by" might indicate that this is so. Either way, the truth-theory would be viciously circular.)

The second way in which our understanding of satisfaction would defeat the theory is if it turns out that we understand the notions by some more direct connection to the schemata, in such a way that all that can be said by way of explaining the notions is simply an appeal to the schemata. This is problematic because if such appeal to schemata is allowed, why should not the truth-theory itself be precisely such an appeal to truth-schemata such as (DS)? That would mean a considerable gain in generality, since sentences of any syntactic structure would be covered by such a theory. This last idea is reminiscent of the critique of correspondence theories in 1.3, according to which the only acceptable elucidation of the notion appealed to in the truth-theory makes the theory unnecessarily complex and incorrectly implying an explanatory dependence where there is none. For both theories, it is argued that *if* only a deflationary account (one appealing to schemata) of the notions in the *analysans* is possible, the truth-theory is inadequate. If only such an account is possible, further, then Tarski's theory suffers, like traditional

correspondence theories, from the deficit of avoiding schemas or trivial equivalences only to find them appear at a later stage. By contrast, deflationism takes such an account to explain truth directly, thereby also giving an explanation why the notions with which truth is explained in these rival theories end up being explained by schemata or deflationary equivalences. This suggests that the only possible alternative for Tarskian theories to be completed is by an independent theory of the representational notions which explains them in some more substantial way than the deflationary accounts.

This leaves the Tarskian idea with a programme for explaining such intriguing notions as that of meaning and linguistic representation, and how truth is related to them. It may be interesting to see why this programme is implied. Hartry Field famously argued that for a Tarskian truth-theory to make truth physicalistically respectable, as he took Tarski to have wanted (1972), the representational notions must be physicalistically reduced. This is here to mean that they are explained by a claim of the form "x satisfies y iff $R(x, y)$", where the relation R is phrased in entirely physicalistic terms, and similarly for denotation. The restriction to physicalistic vocabulary is now generally considered too severe, and one that was not, *pace* Field, required by Tarski (cf. Kirkham (1992: 6.7)). But some other non-circular account does seem required in view of the above considerations. In other words, one need not assume physicalism to see that there is something unacceptable with the list-like account of satisfaction if the definition is to count as a truth-theory (see McGee (2005:115)). Further, though the account need not be purely physicalistic, it seems, as argued by Lewy (1947), that it should come out as a contingent, *a posteriori* matter whether the relations hold between given expressions and objects. That is, it should not be a matter of stipulation, which would be the case if one opted for the first, list-like, type of treatment of the notions (and possibly also on the second one, which uses paradigmatic examples).[6]

It could be objected to this last requirement that David Lewis has shown that there is an admissible construal of these semantic relations as holding of necessity and *a priori* between expressions and objects (1981). This is misleading, however, for on Lewis's construal, what is necessary and *a priori* is that certain expressions are satisfied by, or denotes, given objects *relative to*

---

[6] Peter Pagin has proposed that when Tarskian truth-definitions are taken as truth-theories, they should be held to be contingent. However, I think it intuitively implausible that a claim aspiring to say what it is to be true can fail to be necessary.

a language. Such claims are to be taken as partial definitions of the language in question, and languages, on that terminology, are not necessarily spoken by anyone, but abstract objects, individuated and identified by stipulations to the effect that the expressions therein have certain semantic properties, as specified by the definitions. He does *not* mean that expressions have their semantic properties (i.e., being satisfied by, or denoting, certain objects) necessarily and *a priori*. I may likewise partially define a "person" P as by definition having a sister, so that it is necessary that P has a sister, without any person in the ordinary sense necessarily having a sister. On his terminology, the empirical question about what satisfies English predicates can be phrased as the question of what "language" we speak, in the alternative sense of "language".

Clearly, the criterion of material adequacy as originally formulated cannot be satisfied if the representational relations hold contingently between objects, for that criterion was that the T-sentences are entailed from the truth-definition. It therefore has to be reformulated as requiring that the T-sentences follow from the truth-theory together with the theory of satisfaction and denotation, and the empirical facts about what satisfies and denotes what in various languages. This is a common type of modern correspondence theory of truth, with proponents such as Field (1972) and Devitt (1984: 3.3), and thus one of the main rivals to deflationism. One of the obstacles that must be handled for such a theory to work is that every type of sentence needs to be given a clause in the truth-definition, including all intuitively non-truth-functional sentence-forms such as "p because q", "x believes that p", and so on. (This commitment it shares with the Davidsonian meaning-theoretical programme.)

There are many true thoughts motivating this programme, I believe, and much which is compatible with deflationary theories. Naturally, a sentence of the form "$F(a)$" is true iff what "a" denotes is in the extension of the predicate "$F(x)$". But this does not indicate that truth is in any way conceptually or explanatorily dependent on these relations. Suppose one takes propositions to be the primary bearers of truth. Then, this equivalence can be said to hold because of the way subsentential expressions contribute (contingently, of course) to which proposition is expressed by the sentences in which they occur, together with the obvious truth-conditions of the propositions, given by a deflationary theory. For instance, since "is wise" contingently expresses the property of wisdom, and "Socrates" contingently refers to Socrates, the proposition contingently expressed by "Socrates is wise" is the proposition

that Socrates has the property of wisdom, i.e., that Socrates is wise. By a deflationary theory, this proposition is true iff Socrates is wise. This fact is then generalised for variable sentences of the form "*F(a)*". Since for a sentence to be true is on this hypothesis just for the proposition it expresses to be true, the equivalence above can be shown to be true compatibly with deflationism. One need not agree that a theory of meaning for natural languages must be phrased in exactly this way to see how much of the truth of ideas motivating this meaning-theoretical programme can be true compatibly with deflationism. All that one needs to do, in fact, is to remove the appeal to truth and speak only of sentences' expressing propositions (relative to contexts, naturally). Given a theory which says how the parts of sentences determine what proposition is expressed, the truth-conditions of the sentences (i.e. of the propositions they express) follows trivially given the trivial truth-conditions of the propositions (cf. Soames (1999: 244)).[7] The upshot is that, once again, a true equivalence concerning truth has been conflated with an explanation of truth, just as was argued in the case of correspondence theories. The Tarskian equivalences for particular forms of sentences hold, but this can be given a more natural explanation than that they take part in a theory of truth. If, despite this, one persists in taking a Tarskian approach to natural languages to be a theory of truth, however, quite absurd consequences follow, as we will now witness.

Firstly, one cannot say that Tarski provided a correct truth-theory for certain languages, while taking it to be an open question whether it explains truth for the sentences of natural languages. If a truth-definition really says what it is for the sentences of a particular language to be true (if not, then it is incorrect in any case), then there must be an extension of the definitions that says what it is for a sentence in general to be true (cf. Davidson (1990: 287)). Meeting the criterion of material adequacy cannot plausibly be taken as sufficient for a truth-theory to be correct. A complete and final theory of truth must of course say what makes something true in general, and if the limited

---

[7] Some speak confusingly of propositions *being* truth-conditions. But it cannot denied that propositions have truth-conditions, i.e., that there are conditions under which a particular proposition is true. On the other hand, it may be that all conditions are propositions, for we say, e.g., that a necessary (sufficient) condition for it to be the case that *q* is *that p*. If "that"-clauses always refer to propositions, then all conditions are propositions, including truth-conditions. This last example is more plausibly taken as a way of saying simply that it is the case that *q* only if (if) *p*. Either way, to take the idea that propositions *are* truth-conditions to rebut this line of thought seems confused.

definitions cannot be appropriately subsumed under that theory, it must be incorrect. For example, if the simple truth-definition provided for the two-sentence language above is correct for that language, then the general truth-theory must be an appropriate generalisation of that definition, e.g., the deflationary theory using propositional quantifiers. By "appropriate generalisation", I intend to exclude a theory of truth for variable languages, where every language is given a separate clause in the definition. Doing so is like giving a complete list of red things in order to explain redness. It would mean that in order to understand what it is for a sentence of an unknown language to be true, we need to know about the truth-conditions of certain forms of sentences in that language. As Soames has shown, this is so also if we have a general account of satisfaction and denotation, since explicit mention of the logical constants need to be mentioned in the truth-definition of languages (Soames (1999: 110ff.)). To understand what it is for a sentence of Occitan to be true, one would have to know that if the sentence is of the form "*p* oc *q*", then it is true iff "*p*" and "*q*" are true. This could only be avoided if there were some non-circular way of explaining the notion of "conjunctive sentence", i.e., one which does not explain it in terms of truth. As is easily seen, this type of truth-theory is in general incompatible with truth-theoretic meaning-theories, on pain of circularity. In conclusion, there is reason to believe that the equivalence between "'*F(a)*' is true" and "The object denoted by '*a*' satisfies '*F(x)*'" is yet another example of an equivalence which says something true of "true" without explaining it, just as with epistemic theories of mathematical truth and correspondence theories (see 1.2-3). Consider for an analogy the formulation of a complete grammar for a language which gives necessary and sufficient conditions for a sequence of primitive expressions to be grammatical, that is, a recursive definition. Clearly, such a definition, even if true, does not explain the notion of grammaticality, but is simply a way of listing the syntactic rules for the language. It could equally have consisted in a list of transformation rules. Rather, the term "grammatical" must be presupposed to have a specific content in order for the recursive definition to inform one about the language in question. Similarly, a Tarskian truth-definition can inform us about something, but not so that it explains what it is to be true, but about the language.

The exclusive focus on meeting the criterion of material adequacy in state-of-the-art Tarskian truth-definitions has made them even more ill-suited to give a general truth-theory. These are much more complicated, involving the quite complex notion of denotation relative to a sequence of objects.

Denotation, further, is considered also to hold between variables and sequences. Further, the restrictions upon the sequences relative to which variables are to be considered is also very complex. The problem with all these complications is that they are in no way intuitively related to truth. Rather, they are imposed only in order to meet the criterion of material adequacy. But for the definitions to be acceptable truth-theories, they must all be independently justified, i.e., justified by reference to what seems intuitively essential to truth. Of course, none of the above is to argue that Tarskian truth-definitions are unpromising in other respects, e.g., as meaning-theories that presuppose an independent explanation of truth.

2.5 PROSENTENTIALISM – THE ORIGINAL THEORY

The Prosentential theory of truth, presented by Dorothy Grover, Nuel Belnap, and Joseph Camp (1975) and developed by Robert Brandom (1994: Ch. 5, section III), qualifies as deflationary for this reason: it treats a description of the semantic behaviour of "true" as a full account of truth, no analysis of truth of the (TA)-form is considered, and the truth-predicate is considered as having only an expressive, and no explanatory, role. However, it has been argued by Mark Lance (1997: 183ff.) that one could agree with the prosentential theory insofar as it explains only the functioning of "true", while still believing that more could be said about truth. Thus, one could allegedly be a correspondence theorist about the property of truth while without contradiction maintaining that the prosentential analysis of the functioning of the word "true" is correct. Here, however, I will use the term "Prosententialism" as excluding such a possibility, and follow Grover *et al.* and Brandom in treating it as more of a deflationary theory. We will look in this section at the theory as presented by its inventors, and treat Brandom's views about the matter in the next two sections.

What seems first to have motivated the prosentential theory is an appreciation of the redundancy thesis, as advanced by Ramsey. In particular, their theory is devised in order to vindicate Ramsey's idea that truth ascriptions should be understood as equivalent to certain propositionally quantified sentences, e.g., as with (1) and (1b) above. We saw that one problem about Ramsey's and Williams's theory was that they had no unified theory about "true", but would rather give paraphrases for truth claims as they came,

without trying to say what the analyses had in common. The prosentential theory may be seen as an attempt to rectify this limitation.

To this end, a new grammatical category is distinguished: the *prosentence*. Just as there are pronouns ("she"), proverbs ("do") and proadjectices ("such", "so") (1975: 86), which are expressions referring anaphorically to ante-cedents of the respective grammatical category, there are also prosentences. This term is used by Brentano, who held that "Yes" was a prosentence ("Fürwort") (1904). Another one, suggested by Grover *et al.*, is "so", as in "I don't believe Rachel is sick, but if *so*, she should stay at home" (1975: 88, 91). Grover *et al*. now suggest that the expressions "it is true" and "that is true" should be regarded precisely as prosentences. The alleged difficulty in perceiving this is explained by the syntactic complexity of these expressions. Though apparently of subject-predicate form, these expressions are to be regarded as semantic atoms (tense aside, I suppose). Also, as opposed to "yes" and "so", they can occupy all kinds of sentential positions, just as, e.g., pronouns can occupy all noun positions (1975: 91).

We can now see that some questions concerning (1b), the redundancy theorist's rendering of (1), can be answered. In paraphrasing certain sen-tences in first-order logic into natural language, it is common to use pronouns in order to reveal the structural similarity between the paraphrasing sentence and the one paraphrased. For example, "$(x)(Fx \supset Gx)$" is read: "Everything is such that if *it* is F, then *it* is G." The sentence (1) above cannot be rendered in any such way, since "Everything is such that if he said that it, then it" is ill-formed. What would be needed in order to paraphrase sentences like (1) as we do with first-order sentences would be an expression "which is like a pronoun, but which occupies a sentential position. What is wanted is a *prosentence*" (1975: 82). So, say the prosententialists, with the prosentence "it is true", we can rephrase it: "Everything is such that if he said that it is true, then it is true." (1975: 92). Here, "it is true" is thought to stand to the propositional variable "p" in (1b) in the same way as "it" stands to the first-order variable "x" in the ordinary first-order sentence "$(x)(Fx \supset Gx)$".

Now, one of the important features of these alleged prosentences is their semantic atomicity. To elucidate this idea, a variant of English is imagined where syntactically atomic prosentences are added, "itt" and "thatt", functioning like "it is true" and "that is true", respectively. Thus, "Thatt" would be used just like "That is true", as uttered in response to someone's assertion (1975: 89). More importantly, a quantified sentence like "Every-thing John says is true" could in this language be rendered both as "For every

proposition, if John says that it is true, then it is true" and, by the stipulation, as it were, of "thatt", as "For every proposition if John says that thatt, then thatt" (*ibid.*).

Unfortunately, we are only given the quite rough idea of the semantic rules governing "it is true" that can be sensed by comparison with other proforms, and not much is said explicitly about what they are. There are, however, a few comments made about this, which could be worth mentioning. The prosententialists like to characterise the semantic behaviour of "it is true" in terms of substitution. This is suggested by claims like: "'thatt' is *never* a referring expression" (1975:89 – original emphasis). Also, it is held that "on the face of it, […] questions about what 'thatt' ranges over are misplaced" (1975: 90). Furthermore, instead of speaking of the "range" of quantifiers, it is recommended that we speak of the *instances* of the relevant quantified sentences, where an instance of, e.g., "For each proposition, if John said that it is true, then it is true" is "If John said that Kate is a coward, then Kate is a coward" (1975: 92). Finally, it is explicitly stated that the sentence "Everything is such that if Charley believes that it is true, then it is true" should be understood as "a substitutional quantification, where the truth of [the sentence] is equivalent to the truth of all its substitution-instances" (1975: 113).

Let us look closer at what could be said generally about proforms to see how they could shed light on the semantics of "true", as analyzed by Grover *et al*. What will be relevant for the explanation of any proform is its relation to another expression or class of expressions. Intuitively, a proform stands *for* another expression (from the sense of "pro" in Latin as "in place of"). Such expressions are usually called "anaphoric substituends". To elaborate the intuition, it would seem that a pronoun stands for its anaphoric substituend in the sense that the latter "might just as well" have taken the place of the former (1975: 84). This is the relation commonly called "cross-reference".

However, the above intuitive picture is only correct for certain kinds of uses of pronouns, tellingly referred to as "pronouns of laziness" (after Peter Geach (1967: 627)). Here, the anaphoric substituend is an expression occurring in the same context as the proform itself, and being determined (unless the sentence is ambiguous) by its grammatical type (which should correspond to that of the proform). Such is the "she" in

(G)   If Mary is home from work, then *she* is sick.

or

(G')   *John*: Mary is home from work.
        Bill: Then, *she* must be sick.

Here, one says that "she" cross-refers to its antecedent, "Mary". In these simple cases, the cross-referring relation consists simply in the antecedent's being an expression which "might as well" be in the place of the pronoun, all in line with the common intuitions about pronouns. Now, obviously, the intended sense of "could as well" here is "could without change of what is said". A better phrasing goes by saying that the semantic value of the sentence remains semantically invariant under substitution of the pronoun for its antecedent.[8] This terminology is preferable because it can be extended to other uses of pronouns. We observed, for example, the more important Quantificational Use of Pronouns, for instance (to borrow an example from Grover *et al.* (1975: 85)):

(Q)   Each positive integer is such that if *it* is even, adding 1 to *it* yields an odd number.

In (Q), as Grover *et al.* and many others have noted, we cannot replace "it" by "each positive integer", although the latter seems somehow to be the antecedent of the former. It seems that, rather than being itself an anaphoric substituend of "it", the (restricted) quantifier *determines the class of substituends*. A natural suggestion for explaining (Q) and its likes is to say first that it transposes a positive semantic value (truth, assertibility, or whatever) to *each* instance of "If *t* is even, then adding 1 to *t* yields an odd number", where an instance here is a sentence obtained by replacing "*t*" by a

---

[8] Grover *et al.* object to such an explanation on the grounds that it ignores certain *pragmatic* differences between (G) and the sentence obtained by replacing "she" by "Mary" in (G). In (G), they hold, it is "clear that only *one* person […] is being talked about (1975: 84f.). True, but the ambiguity of the other sentence is surely an effect of the fact that pronouns are used wherever they can. A deviation from this regularity would therefore suggest that there might be two Marys which the speaker ignores to distinguish. In any case, something must account for the fact that "Mary is sick" can be inferred from (G) and its (sentential) antecedent. Also, if we grant a distinction between semantics and pragmatics, the proposed account of cross-reference (where *semantic* value is mentioned) is not touched by the *pragmatic* difference between, e.g., (G) and a sentence where "Mary" substitutes "she".

singular term. Second, since (Q) contains a restricted quantifier, it would also be necessary to restrict further the class substituends of "*t*" as containing only singular terms referring to a positive integer. Alternatively, if one wants to avoid using "refers" in one's semantic explanations, the restriction could be formulated as requiring every substituend *t* to be such that *t*ˆ"is a positive integer" has a specific ("positive") semantic value. Naturally, other variants are possible here[9], but it will do for the moment only to have a rough picture of the kind of explanation that is needed.

The main problem with the prosentential account is that unless we have been told what it is that proforms like "he", "it", "such", "so", etc. have in common with "it/that is true", we have not been told what the semantic functioning of "true" is. It must be said in virtue of what properties it is that "it is true", etc. work in the same way that pronouns, etc., work. What we have been given, instead, is an idle analogy. For all that can be extruded from the prosententialist account is (at most) that the expressions "it/that is true" will *stand to some sentence in some relation similar or identical to that which pronouns bear to their antecedents*. But this is something that everyone (or, at least, every deflationist) agrees with. Since this relation is one of (some kind of) semantic equivalence, we have not been told anything besides the old deflationist claim. The only difference is the unnecessary comparison with proforms.

Note that I am not requiring that the prosententialists give a theory of meaning, viz., that they tell us what specific central notion should be used to account for the cognitive meaning of sentences, be it truth, assertibility, or something else. Whatever the type of semantic value in terms of which the functioning of anaphora should be explained, we still need to be told how "true" affects those semantic values of the sentences in which it occurs.

Now, Grover *et al*. could simply reply here that a sentence like "Everything he says is true" should just be paraphrased into "Everything is such that if he says that it is true, then it is true", and then hold that the role of true here is simply that of allowing the inference to all instances of "If he says that *p*,

---

[9] One variant, of course, might take the notion of reference and/or satisfaction as doing the work, e.g., so that a lazy pronoun is said to take the reference of its antecedent. Incidentally, however, the prosententialists would seem to be committed to a substitutional account of at least some other proforms, since if prosentences (being explained substitutionally) do not behave like other proforms, we do not even have a correct analogy. Thus, since an objectual interpretation of prosentences is counted out, we need to consider only substitutional accounts of other pronouns.

then p", on a par with any other universal sentence. (Or, if we do not wish to use the notion of inference in explanations, we may say that if "Everything he says is true" has a positive semantic value, then so do all instances of "If he says that p, then p".) In both cases, there is a similarity with other quantified sentences, like "Everything is such that if it is blue, then it is coloured", viz., the quantifier "Everything" determines the class of anaphoric substituends. But the problem is that in this explanation, there is no need to mention proforms at all. Thus, there is nothing essentially *proformal* about the word "true". All that needs to be done is to describe in what way the word "true" affects the semantic values of sentences in which it occurs, to specify what inferential relations hold between such sentences and others, and to do this in a finite way that handles all the occurrences of "true". Adding that there are certain similarities between "it is true" and certain proforms is perhaps interesting, but does nothing additional by way of explanation.

It should not be concluded from the foregoing criticism that Prosententialism would be upgraded from the status of analogy only given the correct semantic theory for proforms, or even a whole theory of meaning. On the contrary, given a systematic description of the interrelations between anaphors and their substituends (and probably some more), we could *dispense* with the analogy. For we would then have an account of the way "it is true", etc., function like anaphors, an account of how anaphors work, and, insofar as the account is complete, we would then be able to deduce a description of how "true" affects the meaning of the sentences where it occurs, and there would be no reason to mention any similarity with the functioning of proforms.

I suspect that the mentioning of proforms may have deluded some into thinking that we thereby had a general account of how "true" works. This is incorrect, of course. In fact, there are quite a few types of occurrences of "true", for which the prosentential account gives us no clue. For instance, sentences of the form "It is true that p" and "'p' is true" are just considered equivalent to corresponding "p", but how does this follow from the claim that "it is true" is a prosentence? Thus, there is no unitary account of "true" that enables us in a neat way to derive analyses of various occurrences of "true". This claim will be defended at greater length in the next section, where Brandom's prosentential account will be dealt with. This is for the reason that he goes somewhat further in giving a unified story about "true" as anaphoric. There, we will also see that the prosentential theory is not only inadequate in

its unnecessarily analogical character, but that the analogy in fact breaks down where it seemed to work, e.g., for quantified truth-claims.

2.6 BRANDOM'S PROSENTENTIAL THEORY

Robert Brandom, in his *Making It Explicit* (1994), gives a whole theory of meaning where a vast number of expressions are being explained in accordance with his general non-representational semantics. Two semantically fundamental "statuses" of a speaker are distinguished, "being committed" and "being entitled", and various expressions are then given a semantic explanation in terms of these two notions. In the treatment, much weight is laid on the notion of anaphor, in that many kinds of expression are explained as functioning, one way or other, like anaphors. These, in turn, are given a substitutional explanation. However, partly because the prosentential story about "true" is so short and partly because we are not given any full account of how to determine, for a given anaphor, what its substituends are, it is not possible to infer from these accounts a set of precise statements concerning the semantic properties of "true". We will do our best, however, to reconstruct his claims into something like an exhaustive theory.

There are three types of occurrences of "true" that Brandom mentions and tries to handle. In some cases, his solution is the same as for Grover *et al.*, in others, he goes beyond them. The kinds of occurrences are: (1) *redundant occurrences*: sentences of the form "It is true that *p*", which are intersubstitutable with the corresponding "*p*" in all embeddings (1994: 299f.); (2) *truth-ascriptions to "sentence nominalizations"*, e.g., quote-names of sentences, descriptions like "Goldbach's conjecture" or "'that'-clause sortals such as 'the claim that snow is white" (1994: 300) – here, when, in the first stage, the "sentence nominalizations [have been] discerned, […] a sentence is produced that is nominalized by the locution picked out in the first stage. This is a sentence expressing Goldbach's conjecture, named by the quote-name, one which says *that* snow is white, and so on. It is this sentence that is then treated by theory as intersubstitutable with the truth-attributing sentence, whether occurring embedded or freestanding." (*ibid.*); (3) *quantified sentences*, such as "Everything the policeman said is true", are treated as by Grover *et al.* (1994: 301ff.).

One objection to this account is that it makes different occurrences of "true" function in entirely different ways, and thus fails to meet the constraint of unification. Brandom tries to remedy this flaw by assimilating the explanation of occurrences of type (2) to those of type (3) (though neglecting type (1)). He first considers the strategy of analysing a sentence of type (2), i.e., of the form "*a* is true" as "For any sentence, if it is *a*, then it is true", but finds the analysis grammatically perverse (1994: 304). Instead, he proposes to call the expression "is true" a "prosentence-forming operator", such that when affixed to a sentence nominalisation, it forms a prosentence whose antecedent is the sentence *referred to* by the sentence nominalisation. (1994: 305). For example, "Goldbach's conjecture is true" is a prosentence whose antecedent (for which it is intersubstitutable) is the sentence referred to by "Goldbach's conjecture", viz., "Any even number is the sum of two primes".

The problem here is that, besides the appeal to anaphoric cross-reference, it is difficult to see what occurrences of type (2) and (3) have in common. The former are equivalent to the sentences referred to by the nominalisations, the latter have as consequences instances such as "If the policeman said that *p*, then *p*", but what single fact is it about "true" that explains these two facts? It simply will not do to say that we are dealing with "anaphoric dependencies" in both cases, because in Brandom this expression takes such a wide sense that too much could be explained by such appeal. It is required that the prosententialist say what it is that truth discourse and the use of proforms have in common, that property in virtue of which the analogy is a good one. We said earlier that all that could be extruded from the prosententialist account is (at most) that the expressions "it/that is true" will stand to a sentence in some relation similar or identical to that which pronouns bear to their antecedents. We could thus imagine Brandom trying to make the explanation of occurrences of type (1) into the model by saying that "It is true that snow is white" is a prosentence whose antecedent is whatever comes after "It is true that". Surely, this is correct, on some suitable definition of "prosentence" and "antecedent" (e.g., in terms of preservation of semantic value, as suggested above), but it does no more than repeat what is already by deflationists by using an unnecessary terminology.

But, one might want to reply, if we kept finding matches between various occurrences of "true" and other proforms, this cannot be a mere coincidence. In view of the above, this is not a good argument, but what is worse is that, as will be argued in the remainder of this section, that there are in fact no matches at all. In fact, neither of types (2) and (3) function in a way ana-

logous to any proform. If these criticisms are correct, then there is, of course, something worse about the prosentential theory than its only giving a mere analogy – the analogy does not even hold.

First, let us look at truth-ascriptions to "sentence nominalisations". Here, the whole sentence is regarded as a prosentence whose antecedent is the sentence referred to by the nominalisation. In the trivial sense that the antecedent of a (lazy) proform is an expression that can replace the proform without change of semantic value, the analogy could be said to hold here (at least if we agree with the general deflationist view). For instance, since "Goldbach's conjecture" refers to "Any even number is the sum of two primes", the sentence "Goldbach's conjecture is true" should be at least extensionally equivalent to "Any even number is the sum of two primes". But this is quite uninteresting, given the explanation of "proform" and "antecedent". The problem is that there seems to be no other similarity between sentences of type (2) and any of the traditionally recognised proforms in natural language, that is, no similarity other than the weak one which concerns a certain semantic equivalence between two different expressions. *A fortiori*, there is no other expression in natural language such that the *referent* (in the sense of denotation) of a proper part of that expression is the antecedent of the whole expression. But this was precisely how occurrences of "true" of type (2) were supposed to work. So the analogy fails, and is therefore not only less than an explanation, but directly misleading. (Arguing *ad hominem*, Brandom here also makes use of a non-deflationary notion of reference, contrary to his professed commitment to the impossibility of elucidating such a notion (cf. also Field (2001: 150)). But it is difficult to see how "Goldbach's conjecture is true" could be made sense of on a prosentential account without introducing some such substantial notion of reference.)

We may also note another weakness of the prosentential theory in connection with truth-ascriptions to sentences. This is the impossibility of dealing with sentences of foreign languages. How, for example, should "'Schnee ist weiss' is true" be handled? Brandom says nothing about such cases, but Grover *et al*. suggest that the foreign sentence itself is to be regarded as the substituend of the prosentence. In defence of this daring claim, they appeal to the lack of restrictions in using foreign languages mixed with our own (1975: 102f.). This misses the point, however, since it may be that one understands a sentence like "The sentence '*p*' is true", without understanding "*p*". We could imagine a response suggesting that the account

could be saved by appealing to some notion of translation, so that, e.g., the antecedent of the alleged prosentence "'Schnee ist weiss' is true" is the *translation* of the sentence nominalisation. Obviously, however, this would make the alleged analogy be even more difficult to sustain, for there are hardly any proforms in our language with such a mechanism of translation built into it.

A final complaint about the prosentential treatment of occurrences of type (2) is that it (as disquotational theories) makes the inference from truth-ascriptions to sentences, named by quote-names, to the sentence thus quoted come out valid in virtue of "rules of language". By this phrase, I mean that the inference in question is as strongly valid as that from "If Mary is sick, *she* should stay at home" and "Mary is sick" to "*Mary* should stay at home". However, this is contrary to the intuition that you can infer "Snow is white" from "'Snow is white' is true" only with the further premise that "Snow is white" means that snow is white". This issue will be discussed at length in 2.11.

I said at the end of the foregoing section that there is also a disanalogy even in the case of quantified sentences. One should not be surprised that the prosentential account had *some* appeal despite this alleged discrepancy. For both ordinary quantified sentences of the form "Everything is such that if it is *F*, then it is *G*" and the sentence "Everything is such that if he said that it is true, then it is true" are universal quantifications using the pronoun "it". Thus, the lack of exactness in the use of "proform", etc., in the prosentential analysis makes it easy to see similarities. However, the analysis gains its plausibility from a perceived structural isomorphy that is, as I shall argue, only perceived.

The sentence "Everything blue is coloured" has all instances of the schema "If x is blue, then x is coloured" as instances. But then, on any reasonable grammatical account, the instances of "Everything he said is true" should have as consequences all instances of the schema "If he said x, then x is true". But according to the prosentential analysis, this is a mistake, for it is rather the Ramseyan paraphrase "(p)(If he said that p, then p)" that exhibits the "correct form" of the latter sentence. The reason was that by this claim, they could point at a certain analogy between the following two pairs of sentences:

"Everything he said is true" vis-à-vis "(p)(If he said that p, then p)"

and

> "Everything blue is coloured" vis-à-vis "(x)(If x is blue then x is coloured)".

As said above, the idea here is that "it is true" stands to "p" in the same way that "it" stands to "x". Thus, the prosententialists have to maintain that the Ramseyan paraphrase is somehow the correct analysis of the quantified truth-claim, for otherwise, there would not be enough similarity between "it is true" and "it" to sustain the analogy-claim. But the Ramseyan paraphrase perverts the grammatical form of "Everything he said is true". This will become clearer if we look closer at the pronouns occurring in "Everything is such that if he said that it is true, then it is true". This seems to be a sentence of the form "(x)(If he said that x is true, then x is true)", rather than "(x)(If he said x, then x is true)". Thus, the transformation that the prosententialists make in saying that the real form of "Everything he said is true" is "Everything is such that if he said that it is true, then it is true" (1975: 96), is one where the grammatical structure is changed, and hence, the analogy between "it" and "it is true" to be appealed to holds only for a sentence with a different grammatical structure (namely, "Everything is such that if he said that it is true, then it is true") than the sentence to be analysed (namely, "Everything he said is true").

But why, in that case, did the prosentential analysis seem to work? For the simple reason that the two sentences "Everything is such that if he said that it is true, then it is true" and "Everything is such that if he said it, then it is true" are trivially equivalent. This is for the further reason that: (a) the instances of the variables in term-position (replacing "x") are "that"-clauses, i.e., expressions of the form "that p" and (b) sentences of the form "that p is true" are always trivially equivalent with the corresponding "p". Therefore, an instance of

> (2)   Everything is such that if he said that it is true, then it is true

will always be semantically equivalent to a corresponding instance of

> (3)   Everything is such that if he said it, then it is true.

In virtue of (a), instances of (2) and (3) are exemplified, respectively, by

(2I)   If he said that that snow is white is true, then that snow is white is
       true

and

(3I)   If he said that snow is white, then that snow is white is true.[10]

These two sentences are equivalent in virtue of (b), i.e., because "that snow is white is true" is intersubstitutable with "snow is white". Grammatically, however, it is incorrect to analyse "Everything he says is true" as "Everything is such that if he said that it is true, then it is true, rather than "Everything is such that if he said it, then it is true". In the incorrect *analysans*, the first "it" occurs *within* the primary "that"-clause of "he said", whereas in the correct one, the "it" has such primary "that"-clauses as *substituends*. The prosententialists have thus, as did Williams with his propositionally quantified paraphrases, conflated a certain semantic equivalence with grammatical cotypicality. As the redundancy theorists, they also need to deny the natural idea that "Everything he said is true" is of the same form as other simple universal quantifications, e.g., "Everything he owned is expensive". The plausible deflationary idea that the truth-predicate is introduced in the language with special *semantic* properties in order to achieve expressive strengthening *without* introducing new *grammatical* categories is thus lost. The idea that the claims (a) and (b) show how quantified truth-claims have those semantic properties they have preserves this idea of grammatical conservatism, while also giving a much simpler and clearer account. That these are indeed at the core of a correct deflationary theory of truth will be argued at length in Chapter 4.[11]

---

[10] Concerning the intuition that this is grammatically awkward, see note 4, Chapter 4.

[11] Brandom's (2002), however, might be suggesting that he would agree with me concerning the logical form of quantified truth-claims (see especially p. 107). Here, the role of "is true" in "Everything he said is true" seems to be explained rather by first rephrasing the sentence as "Everything is such that if he said it, then it is true", and then considering "is true" has having the quote-names instantiating the first "it" as antecedent. E.g., in the instance "If he said 'Snow is white', then [it] is true", "it is true" is a lazy prosentence and has the quote-name of the sentence "Snow is white" as antecedent. (This idea also seems easy to extend to cases of indirect speech.) Brandom thus seems to have rid himself of the Ramseyan heritage of the original prosententialists. He also gives a plausible account of tensed and modified truth-claims, so as to save Prosententialism from objections on that score (e.g., in Kirkham (1992:327f.)), and about anaphoric uses of

In summary, the prosentential theory of truth fails for two reasons: first, noticing certain similarities between the expressions "it/that is true" and proforms in natural language, is not yet a full description of the semantic functioning of "true". Further, the treatments of different occurrences of "true" are quite disparate, and so, no unified account has been given. Rather, a full description is given only if we had a full account of proforms and of the connections between them and "true". Given that, however, the mention of proforms would be idle, for we could then derive how "true" affects the semantic content of sentences where it occurs from the account, and so dispense with the comparison with proforms. Finally, it was argued that for two central types of occurrence of "true", there is no analogy with proforms at all.

2.7 BRANDOM ON THE PROPERTY OF TRUTH

It is commonly said that deflationists deny that there is a property of truth. This seems to be taken as a consequence of the deflationary treatment of "true" (cf. 1.4, Thesis V). Brandom follows such a line of thought in taking it to follow from the prosentential analysis of truth-discourse that "is true", despite appearances, is not a predicate (1994: 304f.). Rather, he writes,

> one can see it as a *prosentence-forming operator.* It applies to a term that […] picks out a sentence tokening. It yields a prosentence that has that tokening as its anaphoric antecedent. To take such a line is not to fall back into a subject-predicate picture, for there is all the difference in the world between a prosentence-forming operator and the predicates that form ordinary sentences. (1994: 305, original italics).

The underlying premise is, of course, that only predicates stand for properties. Why, according to Brandom, would it be wrong to grant a property of truth? Brandom first says that such a property is "bound to be 'queer' " (1994: 203). Such an entity, Brandom holds, is only believed by philosophers to exist because they "have misconstrued ordinary talk using 'true' […] on the basis of a mistaken grammatical analogy". By denying that "is

---

"that" in truth-claims (i.e. "That is true"), but both these phenomena will be more clearly explained by the account to be given in 4.3.

true" is a predicate, Brandom can remain a "deflationist" in the special sense that he can deny of "is true" that it corresponds to a property, while consistently affirming it of "has a mass of more than ten grams". This, one might think, would otherwise be impossible, since, as Brandom goes on, "such contrasts seem to presuppose a robust correspondence theory of the contents of some predicates and claims – at least those the semantic deflationist finds unproblematic, paradigmatically those of natural science" (1994: 326). Brandom's point, then, is that by denying that "is true" is a predicate, he can deny that there is a property of truth without having to distinguish between predicates that do and those that do not stand for properties (*ibid.*). It is thus presupposed that distinguishing predicates in this respect can only be done in terms of "robust truth-conditions". The explanation of why there is no property of truth, according to Brandom, is not that sentences containing "is true" do not have "robust" truth-conditions, but, rather, that the expression "is true" "is not even of the right grammatical form to [pick out a property] – anymore than 'no one' is of the right form to pick out an individual, although there are some features of its use that could mislead one on this point" (1994: 327).

The objection against deflationism that Brandom is here trying to respond to is originally Boghossian's (1990). This objection is directed against someone who hold the following pair of views: (1) that "is true" is not used to state facts and does not refer to a property, (2) that a predicate's being used to state facts or referring to a property must be explained in terms of "robust" truth-conditions. The upshot is that *if* the only way to distinguish predicates that stand for properties from those that do not is by claiming that sentences containing the former have robust (correspondence) truth-conditions, while the latter do not, then deflationists can deny that truth is a property only if they have already assumed a "robust" (correspondence) theory of truth, viz., one according to which truth *is* a property. Hence, deflationism is incoherent.

Since Brandom for this reason wants to deny that there is a property of truth, his further view that a predicate necessarily stands for a property forces him to deny that "is true" is a predicate. Let us now focus on his arguments for the view that "is true" is not a predicate. The problem is that of distinguishing grammatical features of "is true" which would mark it off from allegedly real predicates. But questions like these are difficult to discuss without first agreeing on a general account of what a real predicate is. Mark Lance (1997) has investigated this question and his conclusions are quite unambiguous: there is no reason not to call "is true" a predicate, even from

Brandom's own perspective. First of all, he contends, both "is true" and other predicates are "functions from singular terms to sentences" (1997: 190), which is one of Brandom's criteria for being a predicate. Next, Brandom's principal criterion for distinguishing predicates from singular terms, requiring the latter to be governed by symmetric inference-licences, does not seem to give him what he needs. The criterion of symmetric inference-licences unfolds in the following way: take any two sentences *s* and *s'* and expressions *e* and *e'* such that *s* and *s'* differ only in that *s* contains an expression *e*, where *s'* contains *e'*. Now, if *s'* can be inferred from *s*, then *e* and *e'* are singular terms only if *s* can also be inferred from *s'* (1994: 371f.). Predicates (of one place), on the other hand, are functions from singular terms to sentences (so-called "$(T \rightarrow S)$:s" – cf., e.g., Lewis (1972)), and are not governed by this criterion, since they can, but need not, licence symmetric inferences. For instance, "is true" may very well be a predicate according to such a criterion, since "*S* is meaningful" can be inferred from "*S* is true", but not *vice versa* (cf. Lance (1997: 190f.)). Since this is all that Brandom says about predicates in general, nothing in his own account of predicates excludes "is true" from being one, but the most central claims about them makes "is true" a predicate.

Lance also discusses (1997: 191) the possibility of drawing a line between "is true" and other "inferentially asymmetric" $(T \rightarrow S)$:s, and, of course, this could be done in several ways, but none that we have yet seen will disqualify the former as a predicate. In Brandom's (1997) reply to Lance, we are first being given an intricate description of the general semantic features of "is true", given in the prosentential spirit, only to be told, subsequently, that these features "is different in kind from that of picking out properties" (1997: 213). Surprisingly, Brandom's view seems here to be that predicates should be *identified* as the kind of expression which "picks out" (viz., refers to) properties, despite his pronouncements that his semantics is of a non-representational kind. Thus, notions such as reference and truth should not be used as the semantic primitives of his theory. (This argument is, of course, *ad hominem*.) Brandom might reply here that predicates should be identified with those expressions with the (non-representational) semantic features sufficient for "picking out properties". That would be more in line with his general strategy of explaining the representational aspects of language in terms of the practical-social-discursive aspects. However, we are not given any clear description of what these non-representational features are, and the prosentential analysis simply does not, by itself, entail anything about re-

ference to properties. In conclusion, the view that "is true" is not a predicate is currently without support, though much could be held against it.[12]

What now of Boghossian's argument? It may well be that the position taken by him as the deflationary theory (viz., the claim (1), that "true" does not refer to a property) is incompatible with the claim (2), that a predicate's referring to a property must be explained by saying that it forms sentences with robust truth-conditions. However, not only is (1) inessential to the Deflationist position as I have here used this term. Also, the claim (2) is something that deflationists typically deny, both since they would object to there even being such a thing as a robust truth-condition, but also, and more importantly, because they do not think that one should use the notion of truth at all in spelling out the difference between factual and factually defective discourse. Rather, a fully coherent position is that "is true" is a predicate, that there is a property of truth, and that the distinction between "fact-stating" and "non-fact-stating" sentences should not be explicated in terms of truth-conditions. The first of these views is (*pace* Brandom) plausible enough, the second is in no tension with Deflationism (as noted in 1.4), and the third one is commonly seen to follow directly from the Deflationary view. Boghossian's argument is thus directed against a very special position, one, namely, which is the combination of a view not supported by a Deflationary view, namely (1), together with one which goes wholly against it, namely (2). Therefore, it is not one which should worry a deflationist (see Soames (1999: 251ff.) for a more detailed criticism of Boghossian's arguments). Therefore, there is no need, either, to argue that "is true" is not a predicate.

2.8 STRAWSON'S PERFORMATORY THEORY

According to Strawson, an utterance of a truth-ascription amounts to a certain speech-act, that of endorsing, agreeing to, or confirming something. He initially (1949, 1950) went out so strong as to claim that we are not in fact saying anything at all when ascribing truth to a sentence or proposition, but only *doing* something, something on a par with what we do when we utter "I promise that …". He writes:

---

[12] Beebe (2003) also argues for the compatibility between Prosententialism and the existence of a truth-property.

> The sentence "What the policeman said is true" has no use *except* to confirm the policeman's story; but […] the sentence does not say anything further *about* the policeman's story or the sentences he used in telling it. It is a device for confirming the story without telling it again. So, in general, in using such expressions, we are confirming, underwriting, admitting, agreeing with, what someone has said; but (except where we are implicitly making a meta-statement [viz., that so-and-so made a statement], in making which the phrase 'is true' plays no part), we are not making any assertion additional to theirs; and are *never* using 'is true' to talk *about* something which is *what they said*, or the sentences they used in saying it. To complete the analysis, then, of the entire sentence […] "What the policeman said is true", we have to add, to the existential meta-statement, a phrase which is not assertive, but (if I may borrow Mr. Austin's word) performatory. We might, e.g., offer, as a complete analysis of one case, the expression: "The policeman made a statement. I confirm it"; where, in uttering the words "I confirm it", I am not describing something I do, but *doing* something. (1949: 92f.)

An early objection to Strawson's account, raised by Geach (1960: 223) and later Horwich (1998a: 40), is that, since it is an action rather than a claim, a truth-ascription could not stand as a premise in an argument. Strawson (1964) later changed his mind on truth-ascriptions and concurred in Warnock's critique (1964) that it is better to say that we do something (agreeing, etc.) *by* saying something, in uttering truth-ascriptions.

Could it be, then that Strawson means that what one says when uttering, "What the policeman said is true" is the same as what one would have said by "The policeman made a statement. I confirm it." Scott Soames and Jonathan Cohen have made this interpretation and presented a quite persuasive objection. Cohen plausibly argues that this must fail "because [the latter sentence] does not assert a statement which could be verified or falsified by evidence about the policeman's character" (1950: 138). On Soames's more detailed, metalinguistic treatment, the account is held to allow intuitively invalid inferences. Suppose *X* does not know what the policeman has said and yet utter, "What the policeman said is true". Then on the proposed analysis, *X* should be able to infer "There is some statement made by the policeman, which I confirm", but this would be false, since there is no assertion in particular that is being confirmed (1999: 235f.) (cf. Ayer (1963: 165f.)) . Suppose, then, that the only assertion made in an utterance of a truth-ascription is that someone has made such-and-such statements (and

that the "true"-bit is merely performatory). But then, if *X* says something false in situation S, and *Y* utters "Everything *X* said in S was true", then *Y* will have spoke falsely. But the only statement made by *Y*, on this reading of the performatory theory, is that *X* made some assertions in S, which is true (1999: 236f.). Therefore it fails. One could add here that it seems absurd to say that the original pair of sentences, concerning the policeman, could not differ in truth-value: one's confirmation of a statement does not, of course, make it true. Soames also gives a number of sentential contexts in which sentences of the form "The proposition that *p* is true" and "I confirm the proposition that *p*" clearly are not intersubstitutable *salva veritate* (1999: 237f.). One could argue that these arguments are directed against a straw man, however, since Strawson explicitly says that the "I confirm it" part is not to taken to assert something. However, this would mean that the same thing is asserted in the sentences "*x* is true" and "*x* is false", the difference being merely performatory. The rest of this section will show that this proposal meets with grave difficulties.

Many more possible functions than that of expressing agreement, endorsing, confirming, etc., of truth-talk have been observed (e.g., in Kirkham (1992: 310)), and it is reasonable that, with ingenuity, many more could be found. But this seems to be the case with *any* utterance-type. It has to do with the many possible kinds of intentions by which a sentence can be uttered (given some suitable context), and so, there is nothing special about "true" that makes it more of a "pragmatic" expression. If it were, however, it should possible to distinguish some finite amount of types of speech-acts, or some paradigmatic or typical speech-act (as with "I promise…") related to truth-claims, but the number and diversity thereof seems to make this impossible.

It is also debatable whether the alleged principal function – agreeing, endorsing, etc. – is something that can be explained without thereby giving a more direct description of what content truth-ascriptions convey. For in my responding "That is true" to a sentence, e.g., "Snow is white", saying that I am agreeing with the statement seems to boil down to saying that I have myself, albeit somewhat indirectly, said that snow is white. The expression of agreement, by whatever means, is an expression of agreement precisely because it is a matter of saying something with the same consequences and intentions as a mere utterance of the sentence agreed to (with the unimportant exception that by saying "That's true", one has acknowledged the existence of a foregoing assertion). The notions of agreement, etc., are thus explana-

torily superfluous, and may be replaced by some description in terms of "saying something with the same content as", which could further be replaced by a description of the relation of *having* the same content that holds between sentences. But then we are pursuing an ordinary deflationist project, not related to speech-acts at all. There would then be nothing specifically performatory about "true".

Finally, an obvious drawback of Strawson's account is that it cannot properly explain the function of "true" as it occurs in questions, commands, or antecedents of conditionals. According to Strawson, when, e.g., the truth-operator occurs in such contexts, they are there to signal disbelief, doubt or surprise (1964: 78). But, again, these are but a few of the possible implicatures that might come with such an utterance containing "true". Kirkham (1992: 309) notes others, and, again, there seem to be indefinitely many more of them. Given this discrepancy between questions, commands and statements containing "true", the prospect of giving a unified performatory account of "true" seems darker still. For these reasons, Strawson's speech-act account of truth will not be further mentioned in what follows.[13]

2.9 THE LATER WITTGENSTEIN

The comments on truth in Wittgenstein's later writings are quite sparse, but they clearly indicate that he adopted a deflationary view of it. In both the *Philosophical Investigations* (1953: I.136) and in the *Remarks on the Foundations of Mathematics* (1978: Appendix III.6), he cites the schema "'p' is true = p" as explaining the concept of truth. In the *Notebooks 1915-1916*, further, he says that "'p' is true, says nothing else than p" (1961: 9). Of course, the schema is ill-formed in its having an identity sign between two sentence-positions, but we might simply read it is "means the same as" and regard the left-hand side as a variable inside a concatenation-functor and the

---

[13] A somewhat similar idea, inspired by evolutionary biology, is put forward by Huw Price (1988), where ascriptions of truth and falsity function only as "incentives", a behavioural trait which allegedly brings success in that it makes us base our mental attitudes on as wide a body of experience as possible (1988: 150). Again, we may wonder why we shouldn't say that the truth-ascriptions work as such incentives in virtue of what is said by uttering them. Also, all the above arguments against Strawson seem just as cogent against this idea.

right-hand side as just a variable, where the variables range over sentences. In defence of such an interpretation, we could appeal to the passage "The proposition 'It is true that this follows from that' means simply: this follows from that." (1978: I.6).

What is more interesting is that Wittgenstein seems to have been the first to acknowledge an important case of incompatibility between deflationism and the possibility of using the concept of truth in explaining a philosophically central notion. In the *Philosophical Investigations*, the citing of the truth-schema occurs in a context where Wittgenstein criticises various attempts at saying in general what a proposition is (1953: I.134-6). We should add here that Wittgenstein uses "proposition" in an ambiguous way, and it is hard to tell whether he means "sentence" or "proposition", in the now common sense.

The point of Wittgenstein's remarks is that one cannot clarify the notion of a proposition by saying that "a proposition is whatever can be true or false" (1953: I.136). The mistake in such a thought is that

> [i]t is as if one were to say "The king in chess is *the* piece that one can check."
> But this can mean no more than that in our game of chess we only check the
> king. Just as the proposition, that only a *proposition* can be true or false can
> say no more than that we only predicate "true" and "false" of what we call a
> proposition. […]

I take this to mean that truth cannot be used to define a proposition. This would be for the reason that the fact that we only say of propositions that they are true is just a fact about the proper use of the truth-predicate. It would be like saying that the notion of addition must be used in order to define the notion of number, while in fact addition is just defined for (pairs of) numbers, so the notion of number must be understood before we introduce the operation of addition. In a similar vein, Baker and Hacker interpret this passage as meaning that we could not determine whether something is a proposition by *testing* whether it makes sense to apply the truth-predicate to it. For though the resulting expression will make sense only if it is indeed a proposition, the recognition of this depends on a prior recognition of the expression as a proposition, together with the knowledge that "is true" can only intelligibly be applied to propositions (1980: 569ff.).

A similar interpretation could be used to understand Wittgenstein's claim that

[…] the reason why [logical inferences] are not brought in question is not that they 'certainly correspond to the truth' – or something of the sort, – no, it is just this that is called 'thinking', 'speaking', inferring', 'arguing'. There is not any question at all here of some correspondence between what is said and reality; rather is logic *antecedent* to any such correspondence; in the same sense, that is, as that in which the establishment of a method of measurement is *antecedent* to the correctness or incorrectness of a statement of length. (1978: I.156 – original emphasis)

This could be interpreted as meaning that a deflationist cannot use the notion of truth to explain validity of inference (cf. Horwich (1998a: 71ff.)). Wittgenstein may here have meant that since the relation between sentences that we call validity of inference is a meaning-related relation, and since meaning should be explain use-theoretically, so should the former. Wittgenstein's claim of "antecedence" could then be taken as the claim that it must be facts about our practices of inferring, etc., that makes a type of inference valid (as opposed to some notion of truth-preservation). The idea that inference is necessarily truth-preserving would then be explained by reference to these practices together with the deflationary account of truth. More precisely, given that our practices somehow makes the validity from *s* to *s'* valid, the idea that the step is truth-preserving is explained by appeal to the truth-schema, yielding the consequence that we tend to, or should, accept that if *s* is true, then *s'* is true.

These interpretations of the passages would make both of them say quite similar things, namely that substantial facts about meaning and logic cannot be explained in terms of truth, but that the relation between them and truth is simple and can be given by the truth-schema, and that the more substantial explanation of these facts must be given a use-theoretic account. Such an interpretation of the latter passage assumes that Wittgenstein had a deflationary idea of truth in mind in writing it. Since he did not spend too much time in elaborating a deflationary idea, this may seem an unwarranted assumption. Still, if Wittgenstein should be considered an unambiguous deflationist, the above interpretation makes good sense. On this interpretation, further, Wittgenstein would have anticipated Dummett's widely accepted claim that a deflationist must not explain meaning in terms of truth, but must explain it in terms of assertibility-conditions, or, more generally, in terms of the (correct) use of linguistic expressions (1959: 7).

2.10 DISQUOTATIONALISM – QUINE AND FIELD

As I will use the term, "disquotationalism" refers simply to any deflationary theory which takes truth-ascriptions to quote-names of sentences to be primary. Although it seems that we ascribe truth primarily to propositions in ordinary speech, some philosophers refuse to commit themselves to such entities, usually on the grounds that they are thought to be abstract and/or intensional entities. W. V. O. Quine pioneered this view, though presented it quite laconically, whereas Hartry Field has modified and defended it in several papers. I will first present their contributions one at a time, and then present criticisms to the general idea in the next section. An important claim of this book is that a deflationist must take propositions as primary truth-bearers. Therefore, the critique of disquotationalism in this section forms a crucial step in the overall case for the theory I will later defend. The way sentence-truth is to be understood according to that theory is presented in 4.4.

Disquotationalists of course appeal to (DS), or some elaborated variant thereof, in explaining truth. The right-hand side of this schema is what you get if you erase the quotation marks and the truth-predicate of the left-hand side – hence "disquotation". The term is originally associated with Quine (1970), who unambiguously[14] endorsed a deflationary view, as, e.g., the following passages show:

> To say that the statement "Brutus killed Caesar" is true […] is in effect simply to say that Brutus killed Caesar […] (1960: 24)

> Truth is disquotation (1987: 213; 1990: 80)

> By calling the sentence ["Snow is white"] true, we call snow white (1970: 12)

Also, Quine criticizes correspondence theories, firstly, for postulating facts for no other reason than to have something for the true sentences to correspond with (1987: 213; 1990: 80). Secondly, he claims that the theory does not significantly differ from the disquotational theory, since a sentence "It is a fact that *p*" is just equivalent to the corresponding "*p*" (1990: 80).

---

[14] Davidson (1994) has argued that Quine cannot be a deflationist, since he has at various places contended that learning sentences is a matter of learning truth-conditions, which allegedly conflicts with deflationism. I think it would be wiser simply to say that Quine potentially says incompatible things in this case, rather than to regard him as unambiguously a non-deflationist.

Third, Quine's somewhat enigmatic phrase, "Truth is immanent" (1981: 21f.) to a theory or language, seems to be a kind of denial of the correspondence theory. On a simple reading, this phrase can be seen as following from the disquotational view in that a truth-ascription to a sentence not understood will not itself be understood (since a sentence "'p' is true" will have the same meaning for someone as the corresponding "p"). This interpretation may be seen as supported by a definition of "immanence" he gives elsewhere (1970: 19f.), though in connection to grammatical categories:

> A notion is immanent when defined for a particular language; transcendent when directed to languages in general. […] An example of […] an *immanent* notion, is the notion of *der*-words in German grammar. This is a class of words which have the peculiarity of requiring so-called "weak inflection" of a following adjective. It would be silly to wonder regarding some other language, as yet unspecified, what its *der*-words are going to turn out to be. (1970: 19)

So, if truth is immanent in this sense, it would be meaningless to say of a sentence in another language that it was true. But this seems to be on the same footing as just requiring that "true" be applied only to sentences understood. This consequence of disquotationalism will be discussed in 2.11.

As against the above interpretation, the full phrase "Truth is immanent" occurs in such a context (1981: 15-23) that it seems more adequate to regard it rather as an expression of the idea that there is no other way to judge the truth of a statement than by looking at its role in a theory/language and its relation to one's sensory experiences – it cannot somehow be compared to some extraneous world besides this. Or, in Michael D. Resnik's phrase, "we cannot shed our beliefs or language in order to make an unbiased examination of their correspondence with reality" (1990: 405). On any of these interpretations, anyway, Quine's statements are deflationary in spirit.

Quine is commonly credited with first having explained why the truth-predicate, despite its purely disquotational character, is useful to have in a language:

> We can generalize on 'Tom is mortal', 'Dick is mortal', and so on, without talking of truth or of sentences; we can say 'All men are mortal'. […] When on the other hand we want to generalize on 'Tom is mortal or Tom is not mortal', 'Snow is white or snow is not white', and so on, we ascend to talk of truth and of sentences, saying 'Every sentence of the form 'p or not p' is true'

[…]. What prompts this semantic ascent is not that 'Tom is mortal or Tom is not mortal' is somehow about sentences while 'Tom is mortal' […] [is] about Tom. […] We ascend only because of the oblique way in which the instances over which we are generalizing are related to one another. (1970: 11)

This is much in line with the claim of thesis (VI) about the increased expressive power that truth-talk enables.

An immediate worry here is that the truth-predicate would not be fit to serve this purpose unless the inferences from sentences "'*p*' is true" to "*p*" are *direct*. This must be so if the sentence "Every sentence of the form '*p* or not *p*' is true" is to stand to sentences of the form "*p* or not *p*" as a universally quantified sentence stands to its instances. Intuitively, however, they are not direct, but need further premises about the meaning of the sentence said to be true. The idea that whether a sentence is true depends on what it means, Quine himself has called "obvious" (1951: 36). Still, he seems to consider his idea about "true" as merely a device for semantic ascent in conjunction with his scepticism towards propositions and meaning attributions (1970: 1ff.) to enforce the view that any sentence "*p*" follows without the need of further premises (about meaning, etc.) from the corresponding sentence "'*p*' is true". It is telling how Quine suppresses the fact that the truth-ascriptions mention linguistic expressions in phrases like "the truth predicate serves, as it were, to point through the sentence to the reality" (1970: 11), "The truth predicate is a reminder that, despite a technical ascent to talk of sentences, our eye is on the world." and "By calling the sentence ["Snow is white"] true, we call snow white." (1970: 12).

As will emerge, it is important to distinguish descriptive and stipulative disquotational theories. Quine's formulations, however, strongly suggest that he intends a descriptive reading. What I have in mind is particularly his use of "is" in the two first passages quoted above. On the other hand, Quine notoriously chooses laconic, stylistically agreeable formulations over more sober and fully articulate ones, wherefore this interpretation may be gratuitous.

Hartry Field, by contrast, quite explicitly speaks of his truth-theory as a suggestion for how to define a predicate with which the type of expressive strengthening Quine speaks of can be achieved, and expresses doubts about the very idea of distinguishing descriptive from stipulative accounts of concepts (2001: 143). This is a consequence of his firm commitment to Quinean scepticism about meaning (see Quine (1960: Ch. 2)), which also

motivates his eschewing of propositions. He also repeatedly compares different truth-theories in terms of their potential usefulness (1986: 62f; 1994a: § 5). As we will see, however, he later on modifies his theory so as to avoid the most counter-intuitive consequences, which he takes to be an attempt to show how the ordinary use of "true" can be explained in terms of the disquotational truth-predicate (1994a: 266f.).

According to an early formulation of his theory, two essential features of the truth-predicate are that:

(a) it is defined only for sentences that one understands

(b) the property of those sentences which it defines is one that a sentence has or fails to have *independently of the way that the sentence is used by speakers* (1986: 58)

The "use-independence" is meant to have the consequence that:

C_1 if we had used the word 'white' differently, 'grass is white' might have been true

is equivalent (if 'true' is used disquotationally) to:

C_2 if we had used the word 'white' differently, grass might have been white. (*ibid.*)

Thus, Field initially welcomes the consequence that sentences such as C_1 and C_2 above are equivalent. This is the notorious feature of "use-independence" that he imposes on his truth-predicate, which is to ensure that a sentence follows directly from the claim that it is true, and the reason for it is precisely the one I attributed to Quine above, the purpose of semantic ascent (see esp. (1994a: 266)).

To further illuminate the special nature of a use-independent truth-predicate, Field proposes that every truth-ascription, "as a heuristic, to motivate the features of pure disquotational truth", can be read: "true-as-I-understand-it" (1994a: 250). The reason for this requirement is that without such a specification, the claim, e.g., that the axioms of Euclidean geometry might not have been true could otherwise be taken to mean that those axioms might have been used differently and so have been false, which is a trivial claim. Field writes: "what we wanted to say […] 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." (1994a: 266).

Field further proposes a number of modifications of his simple truth-theory so as to deal with certain "problems" (however, these are problems only if the theory is taken to account for our ordinary truth-predicate). First, he tries to amend it so that one can meaningfully apply the truth-predicate also to sentences not understood. To this end, he proposes two varieties of "extended disquotational truth", the first of which is such that a sentence not understood is true in this sense iff it is synonymous with a sentence which is understood and which is disquotationally true; the second defines a sentence not understood as true iff there is a good translation to a sentence understood which is disquotationally true (1994a: 272ff.). It is again Field's Quinean scepticism that makes him speak of "good translations" here rather than synonymy. (However, since "translation" is exactly as problematic as "meaning" or "synonymy", this phrasing seems unnecessary – Field could just speak of synonymy and take that relation to be just as "context-sensitive and interest-relative" as he takes "good translation" to be. He seems to be conflating the concept of synonymy with a particular view of synonymy as objective (1994a: 272f.).)

Secondly, the disquotational truth-predicate seems to have awkward consequences for modal sentences, an example of which is the equivalence between C_1 and C_2. More generally, the truth-conditions of a sentence seem to depend on its meaning or use, but this is explicitly contradicted by the original disquotational theory. To remedy this, Field (1994a: § 9) gives a definition of "quasi-disquotational" truth:

$\Box$ (S is true$_{qd}$ iff $\Sigma p(\exists m(m$ is the meaning of S and $@(m$ is the meaning of "p")) and p) (1994a: 275),

where "Σp" is an existential propositional quantifier and "$@$" is an "actually"-operator, which "temporarily undoes the effect of the modal operator" (1994a: 276). This truth-theory is not obviously deflationary according to Field, because whether it is depends on whether the notion of meaning is explained in terms of truth or not (1994a: 275ff.). In any case, it does seem to yield the more intuitive consequences concerning modal sentences, in that it makes the truth of a sentence depend on its meaning. Field then goes on to say that ordinary intuitions about the truth-conditions of "Snow is white" in circumstances in which it is used in very different ways could be accommodated by saying that the "cash value" of such a claim is just that in the circumstances in which the sentence is used in a different way, it is reasonable to translate it in such a way that its disquotational truth-con-

ditions relative to that translation are, e.g., that grass is red (1994a: 277). He thereby intends to show that the original disquotational theory, which does not endorse the quasi-deflationary definition, is in line with common sense. However, he goes on to say, "I rather doubt that there is a consistent [*sic!*] way to make sense of all ordinary uses of ['true']. I am inclined to think that *many* ordinary uses of 'true' do fit the purely disquotational mould, though I regard the question whether this is so as of only sociological interest." (1994a: 277).

Thirdly, ambiguity and context-sensitivity in sentences seem to yield strange consequences for disquotational theories. An instance of (DS) like "'I am hungry' is true iff I am hungry" seems problematic. Ambiguity also seems to require some modification. In both cases, it seems that the truth-conditions depend on further facts. To justify the intuitions about these cases, Field says that a sentence of either of these kinds is to be taken as true iff there is an association between the sentence and a sentence reading (an "internal analog") free of ambiguity and indexicals, which is disquotationally true. This association is taken to be explained in terms of neural-cognitive processing: "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" (1994a: 280).

It is a little difficult to see what is meant by a sentence reading being disquotationally true, however, since quote-names, for which disquotation is defined, typically refer to ordinary sentences (whether types or tokens). This mystery is somewhat resolved in a different article, where he takes the disquotational truth-predicate to be explained by the claim that "for any utterance u that a speaker X understands, the claim that u is true is cognitively equivalent for X to u itself" (1994b: 405). Further, he says that "for one sentence to be cognitively equivalent to another for a given person is for that person's inferential rules to license [...] the inference from either one to the other" (1994b: 405, n. 1). This "cognitive equivalence", further, is taken to "strictly apply not to the spoken or written sentences, but to internal analogs of them" (*ibid.*).[15] Where it is uncertain which singular term has been taken to

---

[15] Field also suggests (1994a: 259) that substitutional quantification can be employed for further refinements of the theory. However, Field also shows that a formulation in terms of substitutional quantification can be shown equivalent to one where only a notion of inference is used, for instance by "incorporat[ing] schematic letters for sentences into the language, reasoning with them as with variables; and then [...] employ two rules of inference governing them: (i) a rule that allows replacement of

be thus associated in the brain of another, who utters a sentence of the kind in question, "standards of appropriate translation" is the final arbiter (1994a: 281).

In the end, Field concedes, constraints upon the much appealed to "standards of good translation" might have to involve mentioning of non-deflationary semantic features like truth, reference and satisfaction, in order to be adequately determined. But given that a non-deflationary theory of truth could be defined in such terms, deflationism would then fail (1994a: 281).

In general, Field conceives of the rivalry between deflationary and corre-spondence theories in a rather unusual way. In his first article, he speaks of the vindication of either as consisting in a demonstration of the usefulness of either type of truth-predicate (1986: 62f.). Thus, truth-theories are regarded as stipulations of predicates which fail or succeed to the extent that the predicate are useful for given purposes. He conceives of a correspondence theory, further, as one which takes the truth of a sentence to depend on the referential features of the sentence, so that the inference from "s is true" to s requires a premise saying *inter alia* what various terms in s refer to (1986: 60). Elsewhere, he takes a deflationist who takes propositions to be primary truth-bearers to be committed to the view that the expressing relation (bet-ween utterances or sentences and propositions) is insubstantial. Otherwise, he says, we get a substantive, and thus non-deflationary, account of truth for utterances or sentences (1992: 326f.).

Thus, on Field's terminology, a theory which (1) takes propositions to be primary truth bearers and (2) takes the expressing relation between a sentence and a proposition to be substantial, is always inflationist. However, in his (1986), he goes on to discuss the option for deflationists of taking C_1 and C_2 to be inequivalent by taking the truth-conditions of sentences to depend on their meaning, where meaning is to be cashed out in terms of verification-conditions or conceptual role. Obviously, to make this consistent, Field must take verificationist accounts of the expressing relation to be "insubstantial". Probably, "substantial" should here be thought of as related to notions of correspondence. (In the next section, it will be argued that this demarcation of deflationism is incorrect in any case.) He does, however, consider reasons why a correspondence-theoretical truth-predicate might be necessary after all,

---

all instances of a schematic letter by a sentence; (ii) a rule that allows inference of $\forall x(\text{Sentence}(x) \supset A(x))$ from the schema A("p")" (1994a: 259). Since this is correct, I will focus Field's standard type of formulation in what follows.

mainly to account for certain facts about cognitive matters (1986: II-V). In general, Field takes it to be a good heuristic strategy for settling the conflict between deflationary and correspondence theories to assume that a simple disquotational theory is true and investigate whether anything else is needed (1994a: 263, 283f.).

2.11 CRITIQUE OF DISQUOTATIONALISM

We have seen that Field oscillates between two projects in his writings on truth: on the one hand, giving a stipulative definition of a predicate "is true" which is to serve certain theoretical needs and, on the other, giving an account of the truth-predicate as it is actually used in ordinary English. Which of these projects Quine takes himself to be involved in is unclear. Field follows Quine in taking the difference between revising a concept and revising one's beliefs about what falls under it to be a difference in degree only, and consequently considers the distinction between stipulative and descriptive to be "a murky one at best" (2001: 143). Even granted this assumption, however, Field has not given consistent adequacy conditions for his truth theory. For while the earlier Field welcomes many counter-intuitive consequences of his definition on the basis of its alleged "usefulness", and is thus engaged in a "relatively stipulative" account, there is in his later writings an implied commitment to the contrary, since he there takes ordinary intuitions to instruct the formation of his theory, and is thus engaged in a more descriptive account. This critical section will therefore be divided into two parts, which concern the stipulative and descriptive interpretation of disquotational theories, respectively.

If disquotationalism is considered as purely stipulative, one may object, first, that it is misleading to call this predicate "is true", since this easily leads to confusion caused by the discrepancies between it and the truth-predicate we ordinarily use (cf. 1.5). Second, and more importantly, the one purpose the truth-predicate is designed to serve, the expressive strengthening gained by "semantic ascent", seems to be equally well served by some kind of propositional quantification.[16] For instance, Field could just give the rules for a Substitutional Propositional Quantifier thus:

---

[16] Field has granted this in a private conversation.

(SPQ) For any formula ϕ: $\Rightarrow (p)\phi$ iff (for every sentence s, $\Rightarrow \phi(s/p)$),

where the variable "p" is to take only sentence-positions and "$\phi(s/p)$" refers to the expression resulting from the replacement of all free occurrences of "p" by the sentence s. I use "$\Rightarrow$" as a consequence sign instead of the turnstile. It may perhaps be thought that this would not work as intended, because the propositional variables cannot go into quotes. This is easy to remedy, however, simply by stipulating that an occurrence between two quote-marks *is* a sentence-position. Since the quantifier is substitutional, this does not create the problems associated with *objectually* quantifying into quotes. For example, the sentence "(p)("p" is true iff p)" would just *de jure* imply all instances of "'p' is true iff p". Thus, if Field's disquotational truth-predicate is considered as stipulated for the purpose of semantic ascent, it would not only be misleading, but unnecessary.

Let us now consider the disquotational theory as an account of the actual meaning of the truth-predicate. Considering first what Field calls "pure disquotationalism", there is an oft-repeated remark that the instances of (DS), as opposed to those of (PS) and (ES), are contingent and *a posteriori*[17] and, for context-sensitive or ambiguous sentences, odd. The first two facts mean that the left and right hand sides are not intersubstitutable in contexts like "Necessarily/Possibly, p" or "x knows/(justifiably) believes/fears that p", not to mention "mixed contexts", i.e., where contexts of these types are iterated. Theories that take propositions to be primary truth-bearers have much less problems on this score. Indeed, in the next chapter, I argue that the direct inference from "That p is true" and "p" is valid and that they are indeed intersubstitutable in every sentential context *salva* semantic value. In general, there seem to be no true sentences of natural language of the form "Necessarily, if $F(a)$, then $G(b)$", where a and b are terms for distinct physical objects. If one of them is abstract, however, there are, as shown by, "Necessarily, if a is F", then a has F-ness" (where "F-ness" is the relevant term) or "Necessarily, if a has three cars, then the number of cars a has $= 3$" (where "3" is the relevant term).

Another problem is the consequence that truth-ascriptions to sentences not understood are themselves not understood. We normally think that a foreign sentence could be true, for instance if it means that birds fly and birds really

---

[17] Lewy (1947) is the *locus classicus* of this remark, intended to show that Tarski was wrong to make the T-sentences follow from his definition, and so be true by definition.

do fly. On pure disquotational theories, it is only after having learnt what a foreign sentence means do I understand the truth-ascription to it. Thus, if I do not speak any German, I would not understand "'Schnee ist weiss' is true", although I understand both the quote-name (e.g., as a structural description of a sequence of marks) and the truth-predicate (as a device for disquotation) until I have learnt the meaning of "Schnee ist weiss". But if I understand all expressions in a well-formed sentence and recognise the grammatical structure of the sentence, then I should reasonably understand the sentence. In order to avoid this consequence, the treatment of quote-names must entail that a quote-name is understood only if the expression it refers to is, and it is hard to see what such an account would be like.

Of course, Field does not think pure disquotationalism is adequate as an account of our ordinary use of "true". Rather, he attempts, as we have seen, to modify the theory so as to do justice to our intuitions. These modifications are quite problematic. To deal with modal intuitions, Field gives an account ("quasi-disquotationalism") which quantifies over meanings, raising the obvious question why he does not simply take truth to be a property of propositions, and enjoy the benefits of such an account.[18] Ambiguity and indexicality, as we saw above, is handled by taking "internal sentence analogs" to be the bearers of truth. To deal with foreign sentences, the notion of a "good translation" is invoked. But it is quite implausible to hold that ordinary speakers must tacitly operate with such highly theoretical concepts in order to understand "true", as these modified theories imply.

Truth-ascriptions in propositional attitude contexts are not even considered. Nor are the most common truth-ascriptions in ordinary language, where, on the face of it, truth is ascribed to propositions, as in "What he believes is true", and so on. Remember again that modifications for each of these cases is not enough: the theory must also yield correct predictions for sentences where they are mixed. In his excellent *Correspondence and Disquotation* (Ch. 5), Marian David effectively criticizes the modifications actually proposed, even as accounts for the simple cases they are designed for. It seems rather safe to say, then, that if a disquotational theory could be given that, surprisingly, yields the right predictions in all cases, it would be monstrously complex, and to the same extent implausible (cf. Vision (1997: 112ff.)). The alternative which takes (PS) as fundamental, and, as I will argue

---

[18] Volker Halbach (2001), (2002) has proposed various further solutions to the modal problems of pure disquotationalism.

in the next chapter, needs not be supplemented with any claim that is not already independently supported, is clearly to be preferred. This is not surprising, since in all cases that spell trouble for disquotationalism, it is intuitively because sentences express different things, and therefore have different truth-conditions, depending on the context, or because it is what is said by the sentence that seems to have properties like necessity, etc, rather than the sentence itself. I also argue that the preferred theory need not even commit itself to propositions, which is the issue that has motivated disquotationalism all along.

Taking the primary type of truth-ascription to be of the form "That p is true", is thus in better accordance with natural language, and this is of course one of its important advantages. But this advantage is greater than it may first seem. For success in accounting for this functioning entails greater success also in accounting for intuitively non-linguistic facts about truth. For example, when explaining why we are more likely to get what we want if what we believe is true, we can immediately infer the fact without having to say, as disquotationalists would, how the believed things in question are supposed to be related to the sentences that the truth-theory speaks of. Worse still, since disquotationalists typically take sentences as primary because they are sceptical toward the notion of proposition, they have to explain belief as a relation to sentences, a rather unpromising project.

Field often supports his approach in view of this type of objection, with reference to an allegedly chaotic use of "true" and takes the issue to be "of only sociological interest" (1994a: 277). But we have seen in 1.5 that there is good reason to take "true" to be unambiguous, and yet no good reason not to. Ordinary communication involving of "true" simply does not seem to be haunted by any noteworthy risks of misunderstanding, as does communication involving, e.g., "democracy". Field's stance also seems to be in tension with some Gricean version of Ockham's razor, to the effect that one should not multiply senses beyond necessity. It is not in contradiction with this principle, of course, but in the same kind of relation as that between a principle of parsimony concerning, e.g., types of elementary particles and the claim that there are just hopelessly many of them. Further, this issue is hardly of merely sociological interest, since when "true" is used in philosophical texts and seminars, it is very rarely used with reference to some explicitly made definition, but used with a presumption of mutual understanding based on knowledge of its ordinary meaning. If there really were an ambiguity in "true", surely, this would be of great philosophical importance. Buchanan

(2003: 43f.) argues that this is so also if one's only interest lies in arguing replacing the ordinary truth-predicate. Field is not by far the only philosopher who appeals to a presumed linguistic disorder as a smokescreen for empirically inadequate theories. This strategy, together with the appeal to an alleged failure of the analytic-synthetic distinction has become something of an institution.

Finally, I will comment upon Field's view of the demarcation of deflationism. In particular, is it terminologically sound to say that a deflationist who takes propositions to be primary truth-bearers must take the expressing relation to be insubstantial? If he does not, he would indeed have to say that it is a substantial matter that a sentence or utterance is true. But it seems to me that it suffices to be a deflationist to say that it is insubstantial for the primary truth-bearer to be true. The reason is that if one wants an insubstantial account of F-ness, and takes this to be primarily a property of G's, then on Field's criterion, he could not do so unless every relation between G's and a different type of thing is insubstantial. For if H's are sometimes related to G's by substantial relation R, one could define F-ness for H's by saying that an H is F iff it bears R to a G which is F. But then, one cannot be a "deflationist about F-ness", since F-ness for H's would be substantial. Truth can of course derivatively be a property of any kind of thing, as long as it is somehow related to the primary truth-bearers. Whether it is "insubstantially" related to it seems irrelevant to the demarcation of deflationism. Thus, even taking "substantial" to involve correspondence notions, a deflationist should be allowed to hold the expressing relation to be substantial, so that, e.g., what proposition is expressed by a sentence depends on what a name in the sentence refers to. It is incorrect to object, "But then truth for sentences is substantial!". First, the substantiveness does not derive at all from truth. Further, it is misleading to say that, on the propositional account, sentences are derivatively true, because it implies that sentences are true at all. If they are only "derivatively true", they are, strictly speaking, not true at all. Similarly, a visual experience may be said to be red, when it is actually (e.g.) *of* something red. There is no proper sense in which it is red, though certain elliptic phrases may make it seem so. More on sentence-truth in 4.4.

2.12 HORWICH'S "MINIMAL THEORY"

Paul Horwich, arguably the most renowned proponent of deflationism, argues in his *Truth* that all facts about truth can be explained by what he calls the "Minimal Theory" (MT). This is the theory containing as axioms all the instances of the schema (E):

(E) $<p>$ is true iff p,

where "$<p>$" should be read as "the proposition that p". An instance of (E) here is one like $<<$snow is white$>$ is true iff snow is white$>$, i.e., itself a proposition (1998a: 17f.). Horwich admits as a weakness of his theory that it cannot be finitely axiomatised, viz., that it contains infinitely many axioms, although, he argues, an inevitable one (1998a: 25 ff.). I argued in 1.5 that this lack of generality is unacceptable, and will not comment upon it here.

Besides this theory about "truth itself", Horwich also holds that our understanding of the word "true" "consists in the fact that the explanatorily basic regularity in our use of it is the inclination to accept instantiations of the schema (E)" (1998a: 35). It is reasonable that the widespread suspicion against dispositionalist theories of meaning is the reason that Horwich prefers the explication in terms of (MT). Another reason is that he wants a theory of "truth itself" rather than one only of the meaning of "true" (1998a: 36f.). But, as I will argue in 3.5, it is difficult to sustain a theory that gives different accounts of truth itself and the meaning of "true" in this way without entailing claims that go against the deflationary spirit, e.g., that "true" gets its meaning by referring to the property of truth.

Let us focus now on the claim that "(MT) explains all facts about truth". Before the publishing of *Truth*, many philosophers had raised objections of the form "such and such is a fact about truth that deflationism cannot explain". The (alleged) facts mentioned have included the fact that true beliefs facilitate successful behaviour, that meaning is truth-conditions, that it is the truth of theories that accounts for their empirical success, and many more. The method is described in 1.5, where we also saw an example of such an explanation – that of why true beliefs facilitate successful behaviour. Concerning certain theses that are not uncontroversially facts about truth, e.g., that truth is intrinsically valuable, Horwich goes on to argue that *if* it is a fact, then deflationism stands in no tension with it being so (1998a: 62f.). Thus, deflationism is quite neutral concerning this idea of intrinsic value, in

the sense that both proponents and opponents of this idea may both consistently be deflationists (cf. Soames 1999: 231). Whether this is always the case cannot, of course, be regarded as a settled matter, but Horwich's examples seem to indicate that deflationism is at large neutral concerning matters outside the theory of truth (1998a: 57f.).

A well-known problem concerning the explanatory adequacy of (MT) is that general facts seem difficult to explain (cf. Gupta (1993b: 66) and Soames (1999: 247f.)). On (MT), it seems that all one can derive is every instance of a general fact, e.g., the instances of "If the proposition that p and q is true then the proposition that p and the proposition that q are true", but not the general fact itself, i.e., that every true conjunction has true conjuncts. Horwich first proposed that an ω-rule would allow him to derive the general claim, but, as Panu Raatikainen has argued, Horwich seems to be precluded from using the ω-rule, since there are uncountably many propositions, wherefore the rule can never be followed by finite beings (2005: 176). Horwich's latest proposal for dealing with this problem is to add a rule to the effect that "[w]henever someone can establish, for any F, that it is G, *and recognizes that he can do this*, then he will conclude that every F is G" (2002: 68, original emphasis). This was actually designed to explain our acceptance of the general propositions, but the fact-explaining is presumably taken to involve a corresponding claim with a consequent saying that one *may* conclude the universal claim. If this works, then, since the rule does not concern truth, the problem of explaining general facts is solved. We will see in 5.5, however, that there is a different problem concerning infinity that emerges in the context of explaining certain general facts.

Another important feature of Horwich's defence of deflationism is the explanation of semantic contributions of "true" in "blind ascriptions", such as "What/Everything Smith said is true". On Horwich's account, "What Smith said is true" can be explained simply by showing that, given an appropriate instance of (E) as a premise, we can deduce, e.g., "Snow is white" from "What Smith said is true" and "What Smith said = <Snow is white>" (or simply "What Smith said is that snow is white") by simple rules of logic (1998a: 21).[19] A similar account can clearly be given for universal truth-ascriptions. Instead of giving a paraphrase, the functioning of the blind truth-ascription is explained inferentially (though Horwich does not use this expression). We saw in 2.3 that the requirement to give paraphrases lacking

---

[19] A seed to this idea may be found in Ayer (1963: 166).

"true" seemed to require the use of propositional quantifiers, which, however, was shown to yield an implausible grammatical account of the sentences paraphrased. With this inferentialist account, what has been accomplished is a solution of Frege's conundrum discussed above, that the occurrence of "true" in "My conjecture is true", as opposed that in "It is true that snow is white" seems to express something important and irredundant: *we understand ineliminable occurrences of "true" by recourse to eliminable ones*, viz., those of the form "The proposition that *p* is true" (cf. Soames 1999: 230f.).

CHAPTER THREE:
THE PROBLEM OF FORMULATION

3.1 INTRODUCTION

This chapter will be of a slightly more technical character. It will deal with the question of how, in a specific sense, the deflationary theory of truth should be formulated. The question is not that of what is the primary type of truth-bearer (or which type of noun-phrases the truth-predicate primarily applies to), though this is also a major question (and the subject matter of Ch. 4). Rather, the problem here at issue is that of finding an acceptable generalization of the particular fact that it is true that snow is white iff snow is white, i.e., what was in 1.5 called the problem of formulation. In the same section, infinite theories like Horwich's were found untenable on general grounds concerning theory-preference, so we need a solution to this problem if deflationism is to remain a viable theory of truth. I here took the truth-operator "it is true that" as an example, and will continue to do so, but, again, this is an arbitrary choice, since we are not concerned with what truth-bearer or truth-ascription is primary. I take there to be three possible solutions to this problem:

(i)   Schematic solution,
(ii)   Solution by quantification,
(iii) Linguistic solution.

The formulation found mandatory in this chapter, namely (iii) will in the subsequent chapters be argued to yield a theory with considerable merits.[1]

The first type of solution, that we can dismiss right away, is that a sentence-schema like (ES), (PS) or (DS) may itself be the claim with which to explain all facts about truth. Though this has not (to my knowledge) been proposed by anyone as a correct theory, some non-deflationists have spoken as if such a schema is the theory itself. Although we are well accustomed to the use of sentence-schemata in the philosophy of language and logic, we must in this case look more critically to how they are used in order to see why this proposal will not work. To bring the problem to its head, we should note that, strictly speaking, a schema does not *say* anything, since it is not a sentence, but, precisely, a schema. It would not be correct to say that it is *meaningless*, of course, since it obviously partakes in meaningful sentences, sentences which say something. The claim, therefore, is rather that the production of a schema never non-elliptically amounts to an assertion. If it does so elliptically, we had better spell this out and see if the result provides a correct theory of truth. As will be argued below, the possible ways of spelling this out are, besides the above dismissed infinite theories like Horwich's, only (ii) and (iii).

But, it may be countered, in logic one does derive claims from axiom schemata, so why not in a truth-theory? Well, first the practice in formal logic may perhaps be inadmissible, strictly speaking. But we can do better, and note two relevant differences: first, in such theories, the axiom schema can be substituted by the claim that the schema has true instances; second, the axioms do not necessarily have to be claims at all, whereas it seems clearly wrong to say that the theory of truth is not a claim – surely, it is supposed to *tell* us something about truth and *say* what truth is. I conclude that the schematic formulation fails.

Though *prima facie* obvious, the idea that truth can be explained by declaring the schema true has still been questioned, so in 3. 2 I will spend some time refuting it. That refutation will then pave the way for a refutation

---

[1] Discussions about this issue can be found in Tarski (1944), Pap (1952: 8, n. 2), Prior (1971: 22), Mackie (1973: 31, 60f.), Williams (1976: 6-10), Chisholm (1977: 138), Haack (1978: 40, 78f.), Platts (1979: 14f.), Blackburn (1984: 258), Forbes (1986), Baldwin (1989), Resnik (1990: 412f.), Kirkham (1992: 4.4, 5.7), Sosa (1993), David (1994: Ch. 4), Field (1994a: 267f.), Kalderon (1997), Kovach (1997), McGrath (1997b), Horwich (1998a: 25ff.), Soames (1999: 42, 86ff.), McGee (2000), Dodd (2000: 36ff.), van Inwagen (2002), Hill (2002: 17ff.).

of the second type of solution, the solution by quantification. This solution consists in making a claim out of the preferred schema by adding propositional quantifiers. This, one might think, is possible, since the occurrences of "*p*" in (ES) could be made to function as variables. All we need to do, then, is to add a universal quantifier "(*p*)" (obviously not an existential one), and then say that the claim that results is the one which explains truth. As a kind of corollary of the refutation of this solution, deflationary theories of the form "For all *x*, *x* is true iff ...*x*..." which use propositional quantifiers will also be found unacceptable on the same grounds.

The final contender, which will be argued to be the only acceptable one, consists in giving a claim about linguistic expressions, hence "linguistic solution". This may take two forms. In the first, one ascribes some property (other than truth) to the instances of the preferred schema. It will be useful for our coming discussion to see that this kind of solution, where a predicate, call it *F*, is ascribed to the instances of, e.g., (ES), consists in giving a claim equivalent to

(SA1) All sentences *s* are such that "It is true that"ˆ*s*ˆ"iff"ˆ*s* is *F*,

where "ˆ" is the two-place concatenation-function, which, applied to two expressions, refers to the expression obtained by juxtaposing in order the expressions in its two places. (SA1) thus has the form "Every *x* is such that $F(f(f(f(a, x), b), x))$". In the second form of this solution, the claim is that some relation holds between the left- and right-hand sides of the instances of a schema, presumably some kind of equivalence. In analogy with the equivalence between the first form and (SA1), the second kind of linguistic solution, where a relation *R* is said to hold between the left- and right-hand sides of all instances of (ES), will involve giving a claim equivalent to

(SA2) All sentences *s* are such that "It is true that"ˆ*s* bears *R* to *s*,

i.e., a sentence of the form "Every *x* is such that $R(f(a, x), x)$".

On the assumption that these three solutions exhaust the possibilities, it will be concluded that deflationism must be linguistically formulated. How can this assumption be justified? Well, if the criterion on the formulations is so weak as only requiring that they are attempts to generalize over the particular instances, then it seems difficult to give a deductive argument for

it. Indeed, on that criterion, there seem to be an infinite number of ways to form a claim of the kind in question. But let us look closer at the deflationist's basic idea, which is that, e.g., "It is true that snow is white" is in some sense equivalent to "Snow is white", and so on, and that some statement of this fact is explanatorily exhaustive of truth. What is essential here is the structure of the sentences exemplifying the idea, whereas the particular sentence "Snow is white", used in the example, is inessential. It is precisely this structure that is captured by the use of schematic sentence-letters. The point of introducing schemata when reasoning about logic, etc., is to enable us to make generalizations about certain classes of sentences (the instances of the schemata in question). If we are to generalize over the instances without using any additional notions in the generalization, it is difficult to imagine any other possibility than these three. (The solution by quantification may at first seem to go beyond this idea of what is the common denominator of the four solutions, but we will see in section 3.3 that this solution actually turns out to be essentially a disguised form of linguistic formulation on all reasonable interpretations of the quantifier.) Though it does not seem unimaginable that there is some further type of solution, I think it is improbable, and will take the burden of proof to be on the person who thinks there is.

The argument of this chapter, then, aims at the conclusion that a linguistic solution of the problem of formulation is the only acceptable one. This means that a deflationary theory can only be properly formulated as a claim *about the word* "true". More clearly, the only acceptable type of formulation is one where "true" is *not used, but only mentioned*. This was precisely how we regimented Thesis (IV) of 1.4. In a sense, then, this theory is not about "truth itself", only about the word "true". It is important to distinguish this claim from the more common idea that we should focus on the word "true" rather than "truth" in a proper truth-theory. Austin says, famously, "What needs discussing rather [than 'truth'] is the use, or certain uses, of the word 'true'. *In vino*, possibly, '*veritas*', but in a sober symposium '*verum*'" (1950: 117). I agree with what is literally said here, but most, including Austin, seem, by contrasting "truth" with "true" this way, to mean primarily something else, namely, that "true", not "truth", should be used when discussing truth-theories. The conclusion of this chapter, on the contrary, is that none of these words should be used, but that the only acceptable formulation (of deflationism) is one where the latter is *only mentioned*.

Now, as a preparation for the dismissal of the quantificational solution, we need to look closer at the idea that truth could be explained by saying that such and such schema-instances are true. Recall that the problem of formulation was the problem of meeting simultaneously the constraints of finite formulation and non-circularity. The argument for the need to meet the first one was given in 1.5, and the argument concerning this second constraint will be given in the next section.

3.2 EXPLAINING THE SCHEMATA IN TERMS OF TRUTH

As noted, the most natural and common way of understanding the use of sentence-schemata in logic and philosophy of language is to think in terms of their instances being true. This is the reading of schemata generally intended in textbooks of logic, as when the student is asked to say why we should expect "If p, then p" to follow from a good theory, but not "p and not-p". But there is obviously something suspicious about saying that the following sentence (the explication of Truth in terms of Truth):

(TT)   Every instance of "It is true that p iff p" is true

can explain truth itself. (TT), of course, is of the linguistic form that will be argued to be the only acceptable, although (TT) itself, as we will see, fails. Now, although (TT) seems to be using the concept to be explained, we are not dealing with the usual case of circularity, as when a biconditional suffers from too apparent a triviality, as exemplified by an analysis of a concept F of the form "x is F iff x is G", where the synonymy between F and G is too apparent for the analysis to be of any explanatory value.

It may seem that one objection against such a formulation could be given by recourse to the inferential constraint on a successful deflationary theory of truth, i.e., the one which states that the theory must explain in virtue of what the valid inferences depending crucially on truth-ascriptions are valid. We can then follow Horwich in his objection to this type of theory that one cannot infer the instances from the claim that they are true (1998a: 26f.). If so, then one cannot infer the instances of (ES) from (TT), but only the claim that they are true. In order to infer the instances, we would need the instances

of (DS). But, again, we cannot infer these from the claim that they are true. Thus, one would conclude, it is not circularity in its usual form that afflicts (TT), but rather one consisting in some type of presupposition that we are not allowed to make, i.e., that we have explained how to derive sentences from the claim that they are true.

However, one could reply that all that is needed to licence the inferences is to define a valid inference as a Truth-Preserving one, thus:

(TP) $s1, s2, ..., sn \Rightarrow s$ iff$_{def}$ if $s1, s2, ..., sn$ are true, then s is true.

Then, one could assert the instances, because they are unconditionally true. Horwich mentions this reply, but dismisses it as circular (1998: 26). However, I do not see wherein the circularity is supposed to lie, since (TP) is not used to explain (TT), but only to be used in showing how to infer the instances from it.

However, there are other problems with this move. This can be seen by looking closer at the status of (TP). First, if (TP) is regarded as a *stipulative* definition of "$\Rightarrow$", then we do have a problem of showing how to license the inferences, because then, a sentence "A $\Rightarrow$ B" is just an abbreviation of "If A is true, then B is true". Then, (TP) would just be a confusing definition of a symbol which normally has a different use, namely as standing for derivability. More importantly, if (TP) is stipulative, then, with (TT), (TP) and a sentence "A $\Rightarrow$ B", we cannot infer B from A, since the latter just abbreviates "If A is true, then B is true". We can only infer "B is true" from "A is true". As a special case, we cannot infer a sentence A from "A is true". But this is the point of (TP), i.e., of giving a connection between truth and derivability such that with (TP), it can be shown how to infer the instances of a schema from the claim that all instances are true. For this to be made possible, we must thus not *make* (TP) true by stipulation, but *prove* it (and, to be able to do so, assume a prior meaning to "$\Rightarrow$"). Thus, it must be assumed that (TP) expresses, e.g., a material equivalence, a "conceptual analysis", or something else other than a stipulative definition. But if (TT) is to give an *exhaustive* account of truth, then if (TP) is true, it must be proven true on (TT), together with claims not related to truth.

Now, proving (TP) can be done only if one could derive the claim that if A is true, then B is true, on the (hypothetical) assumption that A $\Rightarrow$ B. This, in turn, can be done only if one can derive "B is true" from "A $\Rightarrow$ B" and "A is true" (assuming that "A $\Rightarrow$ B" means "B can be derived from A"). That is,

this is to be done only given (TT) together with non-truth claims. The claim (TT) is the only thing that may be assumed about "true". It is clear that both this, and the converse step, needed to establish (TP), cannot be done. The converse, in fact, is closely related to the very problem we began with, that of deriving a sentence from the claim that the sentence is true. For if it could be shown that B could be derived from A, on the assumption that B is true if A is true, it should also be possible to show the special case in which the set of premisses is empty. That is, it should be possible to show that if A is true, then A can be derived from the null set. This just means that A is categorically assertible.

In response to this, one might, of course, say that (TT) and (TP) *together* are the axioms of the theory of truth. But such a theory would not be deflationary. It would rather be one where it is partly the notion of validity which explains truth. And since it seems hard to find any other way of deriving the instances without adding something like (TP), we have an argument against taking (TT) as the deflationary theory.

Note that this is not the kind of fallacious argument considered by Lewis Carroll, assuming that in order to infer B from A, you always need the premise "If A, then B". If that were right, then you would also need another premise, "If (If A, then B) and A, then B", and so on. I am not claiming that an instance of (DS) cannot be derived from the claim that those instances are true. In a sense, they can, but in order for the theory to meet the inferential constraint, and so be explanatorily exhaustive of truth, the validity of that step must be shown to be valid, in the sense that the conclusion must follow logically from the truth-theory together with the premise (and perhaps some claims not pertaining to truth). If this cannot be done, then the theory fails. If this were not required, then every argument depending for its validity on the truth-predicate would be trivially shown valid, just by assuming it being so. What we want is a unified account of "true" that is to explain why all kinds of inferences that are valid in virtue of it are indeed valid. In this way, it is of course assumed that "true" is less basic than such expressions as conditionals and quantifiers, because it seems that these expressions cannot be given the kind of exhaustive account that we assumed to be available for "true". It seems impossible, for instance, to give a description of the semantics of "Every" which enables us to show why, e.g., universal instantiation is valid, without the explanation involving that very step. It may now be objected that I have assumed without argument that "true" is not as basic as quantifiers, etc. But it is reasonable, for any expression, to prefer an exhaustive account

to a non-exhaustive one, if both alternatives are available. And whereas typically basic logical vocabulary does not seem to allow for such exhaustive accounts, it seems reasonable to believe that "true" does.

Now, surely, there is also something more directly circular about (TT). One case to be made, I believe, is that the circularity consists in the fact that someone not yet knowing what is meant by the claim that the instances of the schema are true cannot be taught that by being given (TT) as an explanation. Such a sentence can equally be taken to implicitly define "false", since substituting "true" with "false" would result in an equally true claim (namely "All instances of "It is false that p iff p" are false"). In order to learn what "true" means from (TT), one must already know what it means, which means that (TT) is not explanatorily exhaustive.[2]

Although one can give a certain piece of information about the semantic properties of expressions in a sentence by saying it is true (as (TT) does), how this is accomplished must be explained by the truth-theory. So, that (TT) gives some information about some semantic properties of the instances must be explained, but cannot be explained solely by (TT) itself. We would thus need a further claim about truth-ascriptions. But this would amount to a deflationary theory only if that further claim was a generalization of the equivalences. So either the appeal to (TT) is idle or one needs to give a non-

---

[2] Ernest Sosa (1993) proposes as a truth-theory

(FMT) For all propositions P, P is necessarily equivalent to the proposition that it is true,

which may at first look like a theory which is both perfectly general and non-linguistic (since "true" is used, not mentioned). However, "equivalence" in (FMT) is explained by

(PE) If <p> entails <q>, then if p, then q.

Now, exactly the problems with taking (ES) to be a theory of truth can now be seen to afflict this combined account. Clearly, if (ES) by itself, or the claim that its instances are true, is not legitimate, neither is (PE). McGrath (1997a) uses Sosa's principles in order to formulate his "weak deflationism" (deflationism about truth of propositions, but not for other bearers) and Kovach (1997) argues against it on grounds of circularity. In reply, McGrath (1997b) gives a putatively non-circular account of "entailment", which includes "For all P, Q, P entails Q iff IF(P, Q) is necessary.", where "IF(P, Q)" refers to the proposition expressed by the conditional from P to Q. However, "necessary" can mean many things ("necessary for survival", etc.). What is intended here is, of course, "necessarily true".

deflationary theory in order to make it work. I conclude that (TT) does not yield an acceptable deflationary theory.

3.3 THE SOLUTION BY QUANTIFICATION

The solution by quantification consists in giving as the fundamental claim of the truth-theory the preferred schema, prefixed by a universal quantifier, as in

(QES) (p)(It is true that p iff p).

This solves the problem that the schemata are not sentences, and so do not say anything. However, there is here a serious problem concerning how to interpret this kind of quantification. We cannot treat the universal quantifier in (QES) as first-order, since (QES) would not be well-formed on such as reading, the last occurrence of "p" standing by its own with no predicate attached to it, and the middle occurrence having instances like, e.g., "It is true that John". Therefore, (QES) cannot simply be taken as intelligible without further comment.

Although philosophers often use propositional quantification quite freely, this cannot be granted the deflationist anymore than the sloppy appeal to schemata discussed above. Our problem is whether we can make sense of (QES) in a way that gives no circularity concerning "true". To judge the viability of the solution by quantification, I will devote this section to examining four kinds of interpretations of the propositional quantifier "(p)": (1) objectual, (2) substitutional, (3) infinitary, and (4) informal understandings of propositional quantification. A final subsection is devoted to truth-analyses of the (TA)-form which use propositional quantification. I will conclude that on all these interpretations, (QES) will not give an adequate account of truth. Therefore, the solution by quantification fails, but it does so in a telling way.

1. OBJECTUAL, SECOND-ORDER QUANTIFICATION

Second-order quantification involves having quantifiers whose variables take positions in formulae other than term position, while still having an objectual

interpretation. The semantics given for such sentences goes by taking the second-order variables to range over objects (hence "objectual"). However, these objects are of a different kind than those over which first-order variables range (wherefore the models in this type of semantics are called "many-sorted"). The objects over which the second-order variables range are themselves assigned extensions in quite the same manner as are the predicates and sentences of first-order logic. These objects are intuitively thought of as properties or classes (for predicate-variables), or propositions or sentences (for propositional variables). Sentences with second-order quantifiers are then taken to be true depending on what is the case concerning such extensions. For instance, a sentence "$(\exists F)(x)(F(x))$" is typically defined as true iff there is a class (in the second order domain), of which every object (in the first-order domain) is a member.

There are now many obvious problems with giving the quantifier in (QES) the objectual interpretation while taking (QES) to be exhaustive of truth. First, the definition is meant to give the truth-conditions of sentences containing the quantifier. The quantifier in (QES) is thus supposed to be explained by the claim that (QES) is true in certain conditions. But if nothing is presupposed about what inferences involving "true" are correct, not much can be done with these truth-conditions. First, from (QES) and the claim giving its truth-conditions, those very truth-conditions cannot themselves be derived, since this would require that the claim that (QES) is true could be derived from (QES).[3]

If, *per impossibile*, this could be done, we would be able to infer that the open sentence (which is identical to (ES)) is satisfied by all objects in the range of the propositional quantifier. So much is required for the account to be objectual at all. But what is it for a sentence or proposition to satisfy (ES)?

---

[3] Need the definition really state truth-conditions for the propositionally quantified sentence? A definition of, say, a predicate, need not give truth-conditions for sentences containing it, so why the quantifier? The answer is that a general account of a propositional quantifier, i.e., which interprets it in any possible occurrence, simply cannot avoid mention of truth-conditions without being schematic. More precisely, it would have to use a schematic letter standing proxy for "open sentences", i.e., sentences with free sentence-variables. But the point of introducing the propositional quantifier was to avoid schemas. If one abandons this generality requirement, one can of course simply take the very sentence (QES) as *definiendum*, and give a *definiens*. But why not take that *definiens* itself to be the truth-theory? This manoeuvre simply consists in letting a certain undefined sentence be the truth-theory, and then defining the sentence by putting it as the left-hand side of a biconditional.

If we assume that the relevant domain consists of sentences, then one is hard pressed to find any other way of defining satisfaction than by the claim that substituting the variable for the sentence yields a true sentence. At least, this should be a consequence of the definition, even if the definition itself may be more complex. It may, for instance, be recursive, giving the satisfaction conditions for every syntactic form of a sentence schema. But each such clause must clearly have the consequence that a sentence satisfies an open sentence iff replacing the variable with the sentence yields a true sentence. So, if the truth-conditions could be derived, we would still only come to know that the instances of (ES) are true. And we already know that this does not suffice for inferring the instances. I doubt taking the domain to consist instead of propositions would be of much help.

To use an objectual semantics to explain (QES) would therefore require that these two problems are somehow solved. But it seems that a semantics along these lines that can also be granted the deflationist would have to be so different from those typically proposed that it seems misleading to call it "objectual" in the first place. All we know about it is that, being "objectual", it is somehow supposed to explain the quantifier by an assignment of some kind of entity to the variables. But given how weak this description is, it seems gratuitous to think that any such semantics will solve the problem of formulation.

There seems to be a more fundamental confusion in the idea that we should look for a way to assign objects to the propositional variables in order to make (QES) a non-circular theory of truth. Assume that the quantifier could be given an appropriately non-circular semantics that assigns a certain kind of objects to the variables. There still seems to be no reason first to formulate a sentence containing the new quantifier, namely (QES), and then give the semantics of the quantifier, so that we can derive something about (ES), instead of directly saying something about (ES) (or about its instances). The quantifier, after all, is to be accounted for by giving truth-conditions of sentences containing it in terms of properties of the rest of the given sentence, i.e., the open sentence following it, which in this case is (ES). (Just as the conjunction is explained by giving truth-conditions of sentences containing it in terms of the rest of the sentence, i.e., the conjuncts.) Being objectual, the account would require a specification of the objects in the range of the quantifier and some notion of satisfaction, as applied to the open sentence. But no matter how this could be accomplished, the resulting interpretation of (QES) could surely be reformulated as a claim about (ES) (or its instances).

Why the indirect manoeuvre? I should stress that this has little to do with the objectual interpretation of first-order quantifiers. If Dummett (1959) was right that deflationary theories exclude a truth-theoretic semantics, then this *account of* first-order quantification is excluded (at least, as an account giving the *meaning* of "all" and "some"). But this is a wholly distinct question. In any case, the deflationist's right to *use* first-order quantifiers is of course not touched by these considerations (pending some surprising argument to the contrary).

2. SUBSTITUTIONAL QUANTIFICATION

This type of interpretation differs from our ordinary, objectual one, in that the truth-conditions of quantified sentences are given by recourse to the truth of sentences in which the variable is substituted for an expression of the appropriate syntactic category. Thus, "$\exists xFx$" is explained as true iff there is a term t such that "Ft" is true, and "$(x)Fx$" is defined as "not-$\exists x$ not-Fx". For any well-formed sentence Φ, furthermore, "not-Φ" is defined as true iff Φ is not true (for a detailed account, see Kripke (1976: 330)). In our case, we would need to speak of substitutions of sentences. In (QES), we would then interpret the quantifier as explained with the addition that the substituting expressions be sentences. But (QES), with the quantifier interpreted in that way, would be trivially equivalent to (TT), which we have already seen to be inadequate. Furthermore, if the quantifier is instead defined substitutionally by giving the conditions for the quantified sentence to be, say, assertible, rather than true, then (QES), together with that interpretation, would just be trivially equivalent to a linguistic solution (of the (SA1)-form), namely, the claim that every instance of (ES) is assertible. The same of course holds for any other property to be ascribed. But that only shows again that the route via (QES) and a definition of the propositional quantifier is idle.

However, there are substitutional interpretations of propositional quantifiers that use the relational property of correct inference, rather than some monadic property like truth or assertibility. Such an account is examined but subsequently rejected by Horwich (1998a: 25f.), but advocated by Baldwin (1989: 101). What seems to be needed here is an Instantiation-rule like:

(I) $(q)...q...$
        _____
         $...p...$

Horwich (1998a: 26) expresses doubts about how this is to be ensured to allow *any* sentence of the form "...*p* ..." to be inferred from "(*q*)...*q*...", but the conventional way of reading symbolisms like (I) of course allow this by *fiat*. To avoid any misunderstanding, let us formulate the Disambiguated Instantiation-rule thus:

(DI)   For any sentence *s* and formula Φ, $(q)\Phi \Rightarrow \Phi(s/q)$.

Here, the variable "*q*" is to take only sentence-position and "$\Phi(s/q)$" refers to the expression resulting from the replacement of all free occurrences of "*q*" by the sentence *s*.

The problem with (QES) on this interpretation is, once again, that it does not yield a claim interestingly different from the claim that the instances of (ES) may unconditionally be inferred (or asserted). One could equally have said that the expression T is a (well-formed) sentence and that every instance of (ES) can be derived from it (and give T as one's truth-theory). One could also, and more naturally, just give a theory of the (SA2)-form and take the relation to be interderivability (i.e., the claim that any sentence "It is true that *p*" and the corresponding "*p*" are mutually inferrable). I conclude that also on the substitutional interpretation, (QES) will either presuppose the notion of truth (such being the common interpretation of quantifiers generally), or (given the inferential definition), it will be equivalent to some form of the linguistic solution.

3. INFINITARY INTERPRETATION

Hartry Field (1984) has proposed that we should understand substitutional quantification as a way of asserting infinite conjunctions or disjunctions. One idea would be to explain the meaning of (QES) by saying that it is equivalent to "(It is true that snow is white iff snow is white) and (It is true that grass is green iff grass is green) and ...". This will not do as it stands, however, since we need to ensure that all and only the intended instances are included in the conjunction. The claim would then have to be that (QES) is equivalent to the conjunction containing all and only instances of a certain schema. But as with the other interpretations, this one is such if it can really give a sensible explanation of the meaning of (QES) (which many will find doubtful), then, if (QES) explains truth, then the infinite conjunction should itself be able to explain truth, which has already been shown wrong. It may look as if we had

a finite claim about truth, namely (QES), but if this is to be explained as equivalent to, or "encoding" (cf. David (1994: 89)) an infinite conjunction, then the infinitary nature of the theory is merely disguised. Again, the proposal is not relevantly different from (but only slightly more complex than) a theory consisting in the sentence "A", which is stipulated to be equivalent to, or abbreviating, or "encoding", the infinite conjunction in question. If the "A-theory" does not solve the formulation problem, it is hard to see why (QES) does. Thus, this idea violates the constraint of finite formulation, though the infinite character is disguised by the finitude of what is taken to be "the theory".

4. INFORMAL UNDERSTANDINGS OF PROPOSITIONAL QUANTIFIERS

It might be insisted that we should not focus on formal accounts of (QES) such as the objectual or substitutional interpretation, but rather on possible informal understandings of the kinds of quantifications needed for (QES), viz., paraphrases into natural language. This does seem possible for other cases of quantification into non-term positions. For instance, it seems that "$(\exists F)(x)(F(x))$" could be paraphrased into "There is something such that everything is it". This is problematic in many ways, but a corresponding paraphrase of propositionally quantified sentences does not even seem possible. As Forbes (1986: 32) notes, contrary to the case of quantification into term- or predicate-position, when we quantify into sentence-positions, there is no copula available to make the claim into a sentence. He writes:

> Quantification into name and predicate position have this in common, that the copula is not absorbed by either kind of quantifier; the special difficulties we are encountering over the interpretation of quantification into sentence position seem to arise because the propositional quantifiers do not attach the variables of quantification to anything that can play an analogous role. (1986: 32)

The best way to clarify this idea, I think, is the following. Paraphrasing ordinary, first-order, quantifications into natural language in the way that most faithfully represents the structure of the sentence (in particular, the variables in all positions and connectives) is by using anaphoric pronouns. By "most faithfully", I mean the way in which there is, more than elsewhere, a one-to-one match between formal symbols and isolated expressions in the

paraphrase. An example is "Everything is such that if it is *F*, then it is *G*" as a paraphrase of "$(x)(F(x) \supset (G(x)))$", which is more faithful than, e.g., "All *F*'s are *G*'s". Now, according to such a scheme for paraphrase, the reading of (QES) into ordinary English would yield the ungrammatical "Everything is such that it is true that it iff it."

It is not inconceivable that resources enabling paraphrases of propositionally quantified sentences could be added to a natural language, however. Arthur Prior envisages such an addition, according to which the sentence "$(\exists p)(p)$" is paraphrased into "For somewhether, thether." (1971: 37). The sentence (QES) would then become "For everywhether, it is true that thether iff thether". The problem with this paraphrase as a contender to an adequate truth-theory is that it is in as much need of explanation as the sentence paraphrased, (QES). I conclude that no informal paraphrase can yield an acceptable deflationary theory.

5. TRUTH-ANALYSES USING PROPOSITIONAL QUANTIFIERS

It has been proposed by a number of authors that one could give a truth-analysis of the form "For all *x*, *x* is true iff ...*x*...", where the *analysans* contains a propositional quantifier. These two variants of such analyses define truth for propositions and sentences, respectively:

(TA3) For all *x*, *x* is true iff $(\exists p)(x$ is the proposition that *p* and *p*)

(TA4) For all *x*, *x* is true iff $(\exists p)(x = "p"$ and *p*).[4]

The question is now whether this use of propositional quantifiers escapes the objections raised against (QES). I argued at the outset that (QES) would need such an explanation, and there is no reason to think that the above analyses would not. The argument, further, against the original proposal was that, given the different ways of explaining the propositional quantifier, the resulting theory would be either circular (by using truth) or just an unnecessarily complex claim equivalent to a claim of the kind involving the

---

[4] (TA3) is discussed in Baldwin (1989), Kalderon (1997: 491), van Inwagen (2002), and Hill (2002: 22), while David (1994: 74ff.) uses (TA4). McGee (2000) discusses analyzing '*x* is true' as 'For all *p*, if $x = "p"$, then *p*'. These variations will not matter to the argument against taking these analyses to be acceptable theories of truth.

linguistic solution. Therefore, the deflationist should opt for the latter kind of solution in the first place. A similar strategy can now be adopted to show that no truth-analysis using propositional quantification will yield an adequate account of truth. First, the definitions in terms of truth are obviously viciously circular in the same way as for (QES). We can also see that what was said about informal paraphrases above shows such an explanation to be impossible here too. The three remaining options are these: either (1) one explains the meaning of a sentence of the form "$(p)\Phi$" by assigning objects to the variables (the objectual interpretation), or (2) a sentence of the form "$(p)\Phi$" is explained in terms of valid inferences involving it (the substitutional-inferential interpretation), or (3) such sentences are considered equivalent to infinite disjunctions of its substitution-instances (the infinitistic interpretation).

Concerning the first proposal, we can see that given that the explication of the propositional quantifier must on pain of circularity be very unlike the known objectual accounts thereof, it again seems gratuitous to believe that some assignment of entities to the variables should somehow give a satisfactory explanation of the *analysans*. Further, imagining such an explanation to be presented, it seems likely that a more direct description of truth-ascriptions (i.e., the *analysandum*) should be possible to extrude from the explanation. In other words, whatever relation (like satisfaction) is supposed to hold between the entities in question and the variables, it seems that if the quantifier is appropriately explained that way, and the *analysans* of (TA3) or (TA4) indeed yields correct results for given truth-ascriptions, then the truth-ascriptions could just as well be explained by appeal to those entities and relations in some more direct manner. And then, the analyses just give an unnecessarily roundabout account.

Similarly, on the second, inferential explanation of the propositional quantifier, it seems that whatever the way the *analysantia* are explained, if the explanation is adequate, and if the analysis is true, then one could instead explain the truth-ascriptions by appeal to these inference-rules appealed to in the explanation of the propositional quantifier. Hence, as with (QES), considering what is required by way of explanation of the propositional quantifier, it becomes apparent that only unnecessary complexity results from explaining truth in terms of propositional quantification, which in turn is explained some way. Finally, the infinitistic interpretation can again be charged with taking an infinite claim to do the explaining of truth-ascriptions while such infinite claims should be proven from the correct theory of truth-ascrip-

tions. For (TA3), the infinite conjunction must be the open sentence "(if $x =$ the proposition that snow is white and snow is white) and ($x =$ the proposition that grass is green and grass is green) and so on". It was argued against taking (QES) together with the infinitistic interpretation of the quantifier as a truth-theory that it was not essentially different from taking a sentence "A" to be the theory of truth, where this sentence is stipulated to be an abbreviation of the infinite conjunction. The proposal was therefore taken to be an infinitary solution in disguise. I see no reason the same could not be said concerning (TA3). The fact that it is an open sentence rather than a sentence proper seems irrelevant, for one could stipulate the open sentence "A(x)" to abbreviate the needed infinite, conjunctive open sentence. That way, the variables would be ensured to be bound by the universal quantifier "For all x" of the analysis. That is, the analysis (TA3) would then read, "For all x, x is true iff A(x)". This does not seem relevantly different from (TA3). But then, (TA3) must be unacceptable, if this analysis is. And, surely, this analysis is also just a disguised infinitary claim.

3.4 THE LINGUISTIC CHARACTER OF DEFLATIONISM

If the reasoning of this chapter so far is correct, the deflationary theory of truth has to be of the form (SA1) or (SA2), i.e., it has to be one about the semantic properties of the truth-predicate, rather than about, in some literal sense, "truth itself". This has been simply assumed by many in the literature, but the crucial argument, I take it, is that the other ways of generalizing the particular equivalences are either unacceptable or only disguisedly non-linguistic. In a sense, Tarski's theory of truth can also be called linguistic, in that it attempts to "give the extension" of the truth-predicate. But whereas it makes sense to say of that theory that it says what it is for something to be true, by giving necessary and sufficient conditions for a variable truth-bearer to be true, this is not so in the case of deflationism. While such truth-theories *can* be formulated so that they *use*, but do not *mention*, the truth-predicate (by giving a standard analysis), a correct deflationary theory cannot. It was suggested that Thesis (IV) of 1.4 should be interpreted precisely as the claim that a deflationary theory cannot be formulated so as to use, rather than men-

tion, "true". Thus, the conclusion of this chapter, again, is that this thesis under that interpretation is true.

This conclusion raises some moot questions that I will discuss in this section. First, it may be thought that the above reasoning, if correct, shows that deflationists are committed to some type of anti-realism. In particular, it may be thought that if the theory needs to be linguistic in the sense made clear, truth will somehow be dependent on language. Presumably, the idea would be that whether something is true depends in general on facts about our language. I do not see how this would follow, however. Whether the proposition that snow is white is true depends wholly on whether snow is white, but whether snow is white need not be taken to be dependent on language, and the conclusion of this chapter does not seem to force the deflationist to say otherwise.

Secondly, it could be argued that the above reasoning simply is a *reductio* of deflationism itself. More specifically, it could be held that the question the theory tried to answer was what truth is, but what is shown here is that a certain answer – deflationism – can only be appropriately formulated as an answer to a distinct question, namely, how the word "true" works, and is therefore inadequate. I believe there are two replies to this natural complaint. First, one can argue that there are other "things" or phenomena that also can only be explained by describing the semantic functioning of the related linguistic expressions, rather than the "things", or phenomena, themselves. Note that it would suffice to persuade the opponent of a single case of this kind to rebut the argument. For the argument, as stated, presupposes that any question about X can only be properly answered by speaking of X rather than the word "X". Without this assumption, the argument is incomplete, since it would require an argument that there is something about truth *in particular* which makes a linguistic theory inadequate. But I know of no such argument that a deflationist could accept.

Some philosophers have found reason to treat such "things" as average persons, sakes and appearances in the same way I propose we treat truth. Perhaps *existence* provides an even better contender. Thus, it could be argued that we should not expect a true and exhaustive theory about existence to give necessary and sufficient conditions for something to exist or in any other way non-linguistically explain "what it is to exist". This notion just does not seem amenable to either form of explanation. Instead, one would argue, the notion of existence cannot be explained "in the material mode" at all, but has to be explained by a linguistic theory about the existential quantifier and/or

the predicate "exists". But if no claim in the material mode does any explanatory work in the explanation of the notion, then it must be concluded that it can be exhaustively explained by a description of the related linguistic expressions, and that there is no meaningful further question about the nature of existence. Perhaps one could construct similar arguments concerning the universal quantifier *vis-à-vis* "universality" or negative existentials/existential quantifications *vis-à-vis* "nothingness". Again, if any of these notions must be thus linguistically explained, the above argument against deflationism fails.

The second reply is that if the deflationary theory of this linguistic kind gives the *meaning* of the word "true", as most deflationists believe, then the *concept* of truth will *ipso facto* have been explained. Next, one could appeal to an intuitive difference between "true" and those words where there is an interesting question beyond the merely conceptual one, as seems to be the case with natural kind terms like "water" and "red". Intuitively, "true" is a non-observational, "logical" expression, and therefore, it could be argued, there is no further question beyond the conceptual one. One can also compare to, e.g., "bachelor" or "doctor", and argue that what it is to be a bachelor or a doctor can be exhaustively explained by describing the relevant concepts, and add that the question of truth is of the same kind.

There is a more specific objection against this kind of deflationary theory, which is that if the theory of truth is linguistic in the sense made clear, then the theory of truth is one about the specifically English word "true", which seems to make it unacceptably provincial. This seems *prima facie* cogent. However, if it could be shown that the theory provides a means for characterizing translinguistically what it is to be a truth-predicate, then, it could be argued, the account is sufficiently general after all. I believe there is such an extrapolation. Suppose that the base-claim the deflationist opts for is that a certain (equivalence) relation E holds between every sentence "That p is true" and the corresponding "p". (I will henceforth take the truth-predicate primarily to apply to "that"-clauses, rather than quote-names of sentences. More on this in Chapter 4.) One could then generalize this idea thus:

(T)   A language L contains a truth-predicate iff L contains a predicate F and a nominalising operator O such that, for any sentence s, O can be applied to s to form a nominalization $O(s)$ and the sentence $F(O(s))$ bears E to s.

The nominalising operator for English is thus the word "that", though a nominal may also be formed by changing the case of nouns and mode of verbs (as in Latin), or simply be identical to the sentence to be nominalised (as I have heard is the case in Chinese). Something like (T) could then be taken as the correct final account, and could be argued to solve the problem of language chauvinism.

Finally, there are serious concerns about the use of semantic notions here, such as the above use of "equivalent", or, for theories of the (SA1) form, the relevant monadic property ("assertible", "analytic", etc.). One argument to be made against this linguistic type of deflationary theory is that the semantic notion used will either presuppose the notion of truth in an inadmissibly circular way, or it will be such that the instances of the schema cannot be inferred.[5] The question is then, first, whether this is so, and, second, whether, if so, this is fatal to deflationism.

In response to the first question, we may note that, assuming that Dummett (1959: 7) was right in deeming deflationism incompatible with taking truth as a central explanatory notion in logic and semantics, the deflationist is, independently of this issue, required to explain the relevant semantic notions by using some other concept. Thus, the possibility for the deflationist of eventually finding the required property stands and falls with the deflationary theory itself (lest we be meaning-sceptics). This means that there can be no objecting to the present use of semantic notions, like correct inference, without, by implication, begging the question against deflationism. The opponent must argue on independent grounds that no other concept than truth can help to explain the notions in question, which is to argue quite a lot.

Suppose, then, that deflationism does not fail in this respect. That means that there is some alternative property that can legitimately be appealed to by the deflationist, which explains such ostensively defined notions such as equivalence, synonymy, derivability, etc. Of course, the relation of equivalence between two sentences must be such that if it holds between two sentences, then the proposition expressed by the first is true iff the proposition expressed by the other is. But this does not mean that the relation must be explained that way. Now, suppose we take the equivalence-relation appealed to by the deflationist to be a very strong one (this assumption will be supported in Chapter 4). Let us call it S-equivalence and define it for expressions of any syntactic category, as follows: expressions *e* and *e'* are S-

---

[5] I owe this point to Paul Horwich.

equivalent iff for any sentence-context $S(\xi)$, $S(e)$ and $S(e')$ are mutually inferrable. The dummy expression "ξ" is to mark the slot in the sentence-context. This slot of course always takes a position corresponding to the syntactic category of the expressions e and e'. If e is a sentence, then the slot takes sentence-position. So, the claim is that for any sentence in which "p" is a subsentence, $S(\text{"}p\text{"})$, the corresponding $S(\text{"that } p \text{ is true"})$ can be inferred from $S(\text{"}p\text{"})$ and *vice versa*. As a special case (taking the sentence-context to be empty), sentences of the form "p" and "That p is true" are themselves mutually inferrable. Given this S-equivalence, the biconditionals (the instances of (PS)) can themselves be derived simply by substituting in sentences of the form "p iff p", which we are independently allowed to assume. Thus, the dilemma is avoided.

Personally, I do not accept the above solution, because I believe the notion of correct inference used in the explication of S-equivalence, and in semantics in general, should be an empirical notion. Thus, it cannot be taken as a regulative rule of inference as such rules are taken in formal systems. This should not be controversial. It is of course a contingent, empirical fact that accounts for the meaning of "true". For instance, we might as well have used "schrue" instead of "true". Whatever it is that gives "true" its meaning, is a contingent, empirical fact. I happen to take it to be the S-equivalence fact just rehearsed. But if it is to be taken as meaning-giving, it must be an empirical notion, i.e., whether S-equivalence holds between two expressions must be an empirical question. Further, if Dummett was right in deeming truth-theoretic semantics incompatible with deflationism, this meaning-giving property must not be explained in terms of truth. But if so, it seems the premise of the objection is true after all, i.e., that the biconditionals themselves cannot be inferred without an illicit appeal to truth. That is, if the notion of inference is not truth-theoretic, then it is not such that the biconditionals can be derived given the S-equivalence. Rather, what can be derived is some "use-theoretic" fact, e.g., that the biconditionals are (categorically) assertible, or "analytic" (where this is not spelled out in terms of truth, but, e.g., as in Quine (1974: 78ff.)). Then, admittedly, the deflationist cannot explain facts about truth in the way this is normally taken to be done, by inferring them from the truth-theory together with other facts.

However, it could be argued that it is sufficient for an adequate theory to explain our "intuitions" concerning truth, in the following sense. If S-equivalence is assumed to hold, then, on a reasonable semantic theory, it should follow that, e.g., the instances of (PS) are assertible (or analytic). Then, it

could be argued, one can explain an intuition about truth if the sentence expressing the intuition (e.g., "If what someone believes is true, then he will be more likely to succeed") could be shown assertible (or analytic). Thus, whereas normally, one takes the task of the truth-theory to be to explain the fact that p by deriving "p" from the truth-theory together with sentences stating other facts, one here instead explains the intuition that p by seeing that the truth-theory implies that "p" is assertible (or analytic), perhaps by being an analytic inference from other assertible (or analytic) sentences. Here, one would presumably have to argue for a close connection between, on the one hand, having the intuition that p" and, on the other, "p" being assertible in one's language or idiolect. Such a linguistic view of intuitions (and related notions like aprioricity) is a well-known philosophical standpoint, and is therefore motivated (whether rightly or not) by other considerations than for the sake of rescuing deflationary theories of truth. Even if many find this view unobvious when stated explicitly, we do not hesitate to accept an explanation to our intuition that, e.g., if blood is red, then it is coloured, on the grounds that it mentions the meaning of the words "red" and "coloured".

This conclusion may appear as a defeat, given that the deflationary theory was supposed to explain the facts about truth, whereas, now, all we can do is derive that various sentences we accept are assertible. But deflationists about truth are typically also deflationists about "fact"; in any case, I am. Such a theory holds, sloppily, that all there is to the notion of fact is the schematic biconditional "That p is a fact iff p". On a deflationary account of "fact", the criterion of exhaustiveness comes to no more than this: for every p, if p, then the theory should entail that p (together with other claims). That is, it is a sentence generalizing over instances of the schema "If p, then the theory should entail that p (together with other claims)". A deflationary theory of facts will of course also have to be linguistic, if the argument of this chapter is correct. From this perspective, there was never any non-linguistic issue to deal with in the first place. Rather, the project of "explaining fact about truth" becomes the project of showing, for every sentence that we accept (containing "true") that our acceptance thereof can be explained by appeal to the S-equivalence and other facts. For instance, we are disposed to accept, e.g., "True beliefs facilitate successful behaviour", and this, we must show, is because of the S-equivalence and a plethora of other things, such as the meaning of other words, past cognition involving relevant terms, perhaps various past perceptions, and so forth. A full explanation of a piece of behaviour is of course a gigantic and rather meaningless project. The important

point is to make plausible that nothing more about "true" is needed in such an explanation. To make it plausible, we can idealize, and simply derive the sentences from the truth-equivalences and other sentences we accept. If all goes well for the deflationist, what will have been achieved is the demonstration that all aspects of our use of "true", *a fortiori*, our accepting various sentences containing it, such as "True beliefs facilitate successful behaviour", requires no more than the deflationist's claim about "true".

We are now in position to close the case opened in 1.1, where the notion of "the correct theory of truth" was introduced. Why, in particular, should we regard the claim which, if taken as primitive, yields the simplest exhaustive overall account, to be the correct theory of truth, i.e., the theory which explains the concept of truth? This claim will be the simplest from which every fact about truth can be inferred. If indeed the S-equivalence involving "true" accounts for all uses thereof in the simplest way, it will also be the simplest way of explaining the "facts about truth", since, for the reasons just given, this comes to no more than explaining our overall use of "true". In the end, I believe, the implicit learning of "true", achieved by treating certain pairs of sentences as equivalent, is a feat of the brain, the emergence of a neural mechanism which, in interplay with other mechanisms, input stimuli, etc., is causally responsible for the overall linguistic behaviour involving "true". If the "explanation of facts about truth" with the S-equivalence works well, then, given the simplicity of the hypothesis, we are justified in holding that the function of the mechanism is to somehow tacitly substitute sentences of the form "That p is true" with "p" and *vice versa*. This is why I take the "objectivist" stance, and treat the question of the correct theory of truth as more than a question of which theory satisfies various desiderata optimally. The factual question is which basic mechanism or "rule" underlies our overall use of "true". The point made at the end of 1.1, that "non-realist" theories of a concept may well be "objectivist" in this sense now becomes relevant: while there is a factual question of the semantics of "true", there is no theory of "what truth really is".

Finally, what can be said about the objection that typically semantic notions involve the notion of truth and therefore make trouble for the deflationist? The "tight connections" between semantic notions and truth are often taken to indicate that the latter must partake in an explanation of the former, and that this is incompatible with deflationism (which I assume to be true). But if this type of argument is sound, we could likewise conclude that truth should be explained in terms of the various semantic notions, wherefore, of

course, it is unsound. The connections pointed out are also easy to explain on a deflationary basis. That propositions have their truth-conditions necessarily follows by a simple intersubstitution of "The proposition that *p* is true" for "*p*" in sentences of the form "Necessarily, *p* iff *p*". The truth-conditions of propositions of course depend on what proposition it is, but so do, e.g., its verification conditions, so that is no reason to believe that the nature of propositions must be explained in terms of truth.

Bar-On *et al* (2000) argue that since the meaning of a sentence determines its truth-conditions, deflationism is either false or compatible with truth-conditional semantics after all. But this connection is also easily explained and does not indicate at all that meaning should be explained in terms of truth, or that truth must be used in a semantic theory. Meaning can plausibly be taken to determine what proposition is expressed by a sentence relative to a context. Taking a sentence to be true iff the proposition it expresses is, it follows trivially that meaning (plus context) determines the conditions in which the sentence is true (i.e., expresses a true proposition), given the intersubstitution claim. This is in no tension with deflationism, since it does not indicate any explanatory dependence. (The converse determination does not seem to hold, however, a notorious problem for truth-conditional semantics.)

Finally, given my (and many other deflationists') reliance on inferentialist explanations, something should be said here about the connection between correct inference and truth. Once again, we should not deny the connection; indeed, if the inference from "*p*" to *"q"* is valid, then if "*p*" is true, then so is *"q"*. This is explained by noting that if the inference is valid, then on any sound semantic theory, "If *p* then *q*" is assertible. If so, it follows by the S-equivalence that "If the proposition that *p* is true then the proposition that *q* is true" is assertible. If we want to show this for sentences rather than propositions, we need only assume instances of "The proposition expressed by '*p*' is that *p*", and again take a sentence to be true if the proposition it expresses is. Gupta (1993b: 79f.) has argued that the truth-predicate must be used in explaining the general fact of which an instance is: "If 'snow is white' can be inferred from 'everything is white', then snow is white if everything is white". Therefore, he says, "the prospects for an inferentialist approach to the meaning of 'true' are bleak indeed" (*ibid.*). Given the above claim about how to explain intuitions of this kind, we should take this instance to be explained by the semantic theory's general consequence that whenever a sentence "*p*"

can be inferred from a sentence "*q*", the sentence "if *q*, then *p*" is assertible (or analytic).

3.5 A FURTHER ARGUMENT FOR THE LINGUISTIC FORMULATION

This section is devoted to giving an additional argument that a linguistic formulation is mandatory for a deflationist, *insofar as* s/he accepts thesis (VI), that the truth-predicate is *merely* a device for increasing the expressive strength of a language. It is crucial that everyone (as far as I know) agrees that "true" does have this function. The important issue is whether it is the *only* function, or if it is a by-product of some underlying property. If this thesis is true, I will argue, one can only explain the equivalence between the sentences "That *p* is true" and "*p*" by saying that they have been conventionally stipulated to be equivalent, rather than being equivalent both because of a conventional stipulation and further facts. This, in turn, can only be done on a linguistic formulation.

The difference between these "ways of being equivalent" can be clearly explained by comparing the following pairs of sentences: "*x* is a bachelor"-"*x* is an unmarried man" and, on the other hand, "*x* is a renate"-"*x* is a cordate" (on Quine's coinage). The latter two sentences are equivalent *both* because of a conventional stipulation and an empirical matter of fact. The conventionality is not meant to come to anything more than the trivial fact that we might as well have used "schmordate" and "schmenate". Clearly, this linguistic convention does not by itself determine the (material) equivalence between the sentences in this case. The former two sentences, by contrast, are not equivalent due to any other fact than a linguistic convention to treat the expressions as synonymous. The word "bachelor" is not introduced by some kind of ostension to various objects, which we may or may not discover to be the same as the things we call "unmarried men". These are in any case assumptions that I find plausible, even if some old-school Quineans may want to disagree.[6]

---

[6] Quine would not – cf. *inter alia* his (1974: 78ff.) and (1991: 270). What *might* be argued, at most, is that he has consistently denied that there is a defensible notion of analyticity with the type of epistemological significance assumed by Carnap and others.

One might want to object that this unduly presupposes that the intro-duction of "bachelor" could not instead be explained as a conventional asso-ciation between it and a *meaning*, which happens to be the very meaning with which "unmarried man" is conventionally associated. But, in fact, I could allow this type of explanation. On such a view, the issue in question is whether "true" is introduced by the convention to associate "that *p* is true" with whatever meaning that is associated with the corresponding sentence "*p*". Such an account will serve the point I will make equally well. On such a view, the question is still whether the equivalence between "That *p* is true" and "*p*" is an effect purely of a conventional stipulation, but the stipulation is somewhat differently accounted for. The crucial common feature is that on this view, one cannot understand the expressions without taking the sentences in question to be equivalent. The same goes for the former account of con-ventional stipulation. Accordingly, on this view, the sentences are not equi-valent because of facts that are independent in any way of what speakers need to know in order to use the expressions competently. The same goes for the former account of conventional stipulation.

The point now comes to this: if "true" is introduced *merely* as a device of expressive strengthening (of the familiar kind), then the deflationary theory to be coupled with this claim must be linguistically formulated. *A fortiori*, it must be the claim that sentences "that *p* is true" and the corresponding "*p*" are equivalent in virtue of a conventional stipulation and no more. I will argue that this cannot be made justice on a merely non-linguistic formulation, since such formulations, by our definition, do not essentially mention the word "true", but *use* it. The remaining option is that of having a non-linguistic theory of truth itself *coupled with* a theory of "true". This, however, can accommodate the idea of expressive strengthening only with an im-plausible consequence concerning the relation between truth and "true". Or so I will argue. The argument is of course distinct from that which has mainly occupied us in this chapter.

Beginning with a non-linguistic formulation, such as the infinite "Minimal Theory" of Horwich, the problem is that of taking this theory alone to be exhaustive of truth and simultaneously hold that "true" is merely a device of expressive strengthening. If such a theory is taken as exhaustive, then there may not be an additional claim about "true" (as there is on Horwich's combined account). The properties of "true" must then be given by an asso-ciation with the property of truth, as characterized by the non-linguistic theory. For instance, the word could be said to refer to the property in ana-

logy with the way other adjectives refer to various properties. This is problematic as it is, but is not my main target here. Rather, the problem is that on this account, the expressive strengthening of "true" cannot be said to be the only function. On this type of account, rather, this function is a by-product of its referring to the property of truth, and the nature of this property. This account does not differ from correspondence theories (and other theories), on which "true" plays the expressive role in question, but is not introduced *merely* in order to play it, as deflationists hold. The equivalence between sentences "That *p* is true" and "*p*", which results in the expressive strengthening, is not here merely conventionally stipulated, but is explained by appeal to the nature of truth, just as the equivalence between "*x* is a renate" and "*x* is a cordate" is explained by appeal to the nature of cordates and renates. Such an account, accordingly, would make speakers' knowledge that the sentences are equivalent dependent not merely on their understanding the word "true" but on this *together* with their knowledge about the property of truth. It would not be able to take the equivalence between the sentences as basic, but would require further facts to explain it.

In the case with "renate" and "cordate", the equivalence holds because of a contingent fact, of course, but I do not see that the modal difference between the cases is relevant. It is relevant only if necessity is treated, in logical positivist fashion, as a linguistic phenomenon. For then, one could say that whereas the equivalence involving "renate" is partly factual while the truth-equivalences are merely linguistic. But on such an account of modality, together with the view that the equivalences are indeed necessary, one has already agreed with a wholly linguistic account of truth. Modality is not at issue, but only whether a certain concept can be explained merely by reference to intralinguistic relations, or if it also requires appeal to extra-linguistic facts. If one *denies* the necessity of the truth-equivalences, on the other hand, one cannot, of course, object to the analogy with "cordate" and "renate" on modal grounds.

The upshot is that a purely non-linguistic theory of truth cannot do justice to the idea that "true" is introduced merely as a device of expressive strengthening, but must agree with correspondence (and other inflationary) theories that this function is a by-product of "true" referring to truth and the nature of the latter. In order to avoid such a non-deflationary consequence, one could perhaps say both that the equivalence is conventionally stipulated, and also that the nature of truth is explained by a non-linguistic theory. Specifically, the nature of truth might be explained by reference to the fact

that the proposition that snow is true iff snow is white, and so on (Horwich's account). This we may call the *dual account*. The problem with this view is that it makes the property of truth and the word "true" implausibly unrelated. The explanation of one does not make any reference to the other. But surely, if there is to be any mention of the property of truth, the explanation of "true" had better be explained by reference to it.[7]

It may be thought that the dual account may be saved by the idea that although truth itself and "true" are explained differently, they are still related in that the latter refers to the former *in virtue of* the conventionally established equivalence. This is somewhat reminiscent of Frege's idea that we grasp the concept of direction by taking as defining of "direction" the sentence "*a* and *b* have the same direction iff they are parallel" (Frege (1884: § 64)). But it is rather mysterious how a convention to treat two sentences as equivalent can enable the grasping of something external that was not previously grasped. This is of course unproblematic if the "concept-grasping" is interpreted simply as the ability to correctly use the expression, and not something resembling Platonist quasi-perception. Thus, if the reference of "true" to the property of truth is taken as a long-winded way of saying that it means *true*, and this in turn is explained *without* reference to the property of truth, i.e., by a purely linguistic theory, then my only disagreement concerns the misleading terminology. But on the dual account here considered, the mention of the property of truth is precisely *not* irredundant in this way, and therefore, if the argument of this section is sound, the account either postulates a mysterious connection, or conflicts with the claim that "true" is merely a device of expressive strengthening.

In summary, it is only a purely linguistic theory which can accommodate the idea that "true" is merely a device of expressive strengthening. This idea is here (and generally) taken to exclude the view that the expressive function is a by-product of "true" referring to truth, together with facts about, or the

---

[7] This disunity of Horwich's combined view is easily overlooked, and probably because of a three-fold ambiguity in the word "equivalence", which may refer to (1) a certain kind of proposition (e.g., the proposition that the proposition that snow is white iff snow is white, (2) a certain kind of sentence (e.g., "The proposition that snow is white is true iff snow is white", or a certain relation between sentences (the one that holds between "That snow is white is true" and "Snow is white"). If one slides between the first two of these, it may seem that the equivalences that explain truth are the same as those the categorical acceptance of which is necessary and sufficient for understanding "true". Once one sees the difference, the question how the two facts are related create the problem here at issue.

nature of, truth itself. Everyone agrees with this claim and therefore is not the deflationary idea of "true" as merely an expressive device. The reason why anything beyond a purely linguistic theory cannot accommodate the idea is, again, that "true" must either be characterized in terms of referring to truth, whence the equivalence between truth-ascriptions and denominalized (or disquoted) sentences is partly explained by the nature of truth, or, that a dual account consisting of a theory of truth proper and a theory about "true", will not make the one relate to the other save by an inexplicable connection. A linguistic formulation, which holds that there is no theory about truth itself, but only about the semantic properties of truth-predicates, clearly avoids the dilemma.

CHAPTER FOUR:
THE PRIMARY TRUTH-ASCRIPTION

4.1 INTRODUCTION

The foregoing chapter left us with the conclusion that the deflationary theory can only be properly formulated as a claim about linguistic expressions, more precisely, as a claim of the form (SA1) or (SA2) (cf. 3.1). As anticipated in 3.4, I will take the basic claim to be of the form (SA2), and to be:

(D)   Every sentence of the form "that p is true" is S-equivalent to the corresponding sentence "p".

To repeat, expressions e and e' are S-equivalent iff for any sentence-context $S(\xi)$, $S(e)$ and $S(e')$ are mutually inferrable. A sentence-context may simply be seen as a function from the expression in question to a sentence containing it. So, the claim (D) makes is that for any sentence in which "p" is a subsentence, $S("p")$, the corresponding $S("that\ p\ is\ true")$ can be inferred from $S("p")$ and *vice versa*. As a special case (taking the sentence-context to be empty), sentences of the form "p" and "That p is true" are themselves mutually inferrable. I will leave it open which view of inference should be adopted. This is a controversial issue, since it seems the deflationist is not allowed to give a truth-theoretic explanation of it. If this is so, and if a truth-theoretic explanation is required, then deflationism fails. I will not discuss this issue here, but, rather, try to show that *if* this notion (and other semantic notions) can be explained in a way compatible with deflationism, then, *whatever this explanation is*, (D) enables us to explain a host of problematic facts about "true". The explanation must of course meet obvious constraints, such as validating the claim that "p" can be inferred from "p and q" (in

English), and so on. Thus, it is to justify our pre-theoretic intuitions about inference. This chapter will be devoted to defending (D), as well as showing how it avoids a number of problems that have afflicted other deflationary theories. These problems are those of:

(1)   explaining how "true" contributes to the semantic meaning of any sentence where it can occur, and doing it in a unified way (cf. the inferential constraint and the constraint of unification in 1.5)

(2)   explaining the problematic intuitions concerning sentence-truth,

(3)   steering between the Scylla of admitting propositions and the Charybdis of disquotationalism (the unacceptability of which was argued in 2.11),

(4)   enabling general facts about truth to be inferred, which is part of the explanatory constraint of 1.5.[1]

In this chapter, I will first defend (D) and then show how (1) and (2) can be accomplished, while (3)-(4) is left to Chapter 5. I should forewarn that I will sometimes use "It is true that p" instead of "That p is true" in what follows, to avoid certain awkward constructions. The switching between these forms will be shown innocent in 4.3, where the "truth-operator" is discussed.

One might say that arguing that (D) is the correct account of "true" is simultaneously to argue that truth is a property primarily of propositions. The conclusion of the previous chapter requires this primacy claim to be metalinguistically formulated, however, as does the nominalist account of propositions that will be given in the next chapter. For simplicity, I will

---

[1] For lack of time and space, I have had to omit a fifth problem here, namely that of explaining how the Liar paradoxes can be given a neat and simple (dis)solution, given a use-theoretic and linguistic account of truth. The idea was that natural languages are inconsistent in virtue of contradicting rules governing "true", negation, and other expressions, and that with the advertised kind of theory, this could itself be consistently and plausibly described (cf. Chihara (1979), (1984), and Burgess (2002)). That seems to be the only way to do justice to the ubiquitous intuition that the derivation of the contradiction, for all its troublesome consequences, is correct. It thus seems to be the only solution that does not violate intuitions, and so, the only one that is not in fact a non-solution.

occasionally switch to "the material mode" in discussing this issue, that is, arguing more straightforwardly that propositions should be considered the primary bearers of truth, rather than discussing what is the primary grammatical object of the truth-predicate. This switch will be discussed and defended in the next chapter.

4.2 A DEFENCE OF INTERSUBSTITUTABILITY

There are first a number of alleged counter-examples to the substitution claim that should be considered. First, in making intersubstitutability-claims, it is commonplace to state exceptions pertaining to quote-contexts. However venerable, I will argue that this is not mandatory, but that there is a view about quote-names with the consequence that this exception need not be added, and which, in fact, is to be preferred. When claiming that an expression *e* of a certain grammatical category can substitute another, *e'*, one does not mean that any string of letters or phonemes which constitutes *e* can substitute any string of letters or phonemes which constitutes *e'*. Borrowing an example from Quine (1953: 140), someone who claims that synonyms are substitutable does not mean that one can substitute in "cattle is expensive" so as to get *"felinetle is expensive", although "cat" and "feline" are synonymous. But surely, the adherent of this substitution view is not required to specify that the string of letters "cat" must occur as a noun in order to be substitutable. What I shall now argue is that what comes inside the quote-marks in a quote-name of a sentence does not occur *as* a sentence anymore than the string of letters "cat" occurs *as* a noun in the word "cattle". This is a consequence of a certain principle of compositionality, namely, that if an expression *e* of category C occurs as an instance of C in a sentence, then the semantic content of *e* will contribute to that of the sentence. But clearly, the semantic content of "Snow is white" does not contribute to that of the sentence

"Snow is white" contains three words.

If a linguistic theory satisfies this principle, I believe, then it is *ceteris paribus* to be preferred over one that does not. If so, quote-names should be seen as referring to a type or token string of letter/phonemes determined by

the marks that occur inside the quote-marks, but that what comes inside the quote-marks is not a sentence. This view is further evidenced by the fact that quote-names of signs that have never been used before (or of ill-formed strings like "kkrq7") are perfectly intelligible, and refer to tokens of the same shape or that type of shape. What occurs in the quote-name above is of course a string of letters belonging to a type of strings some of whose tokens occur as sentence-tokens, but it does not itself so occur. Likewise, the string "cat" occurs in "cattle", but not as a noun, but as a string belonging to a type some of whose tokens occur as nouns. Thus, since what (D) claims is that certain *sentences* can be substituted for others, it need not be supplemented with exceptions for quote-contexts. One may wonder whether this is consistent with the apposition "The *sentence* 'snow is white'". It is. One can (and should) say that quote-names can refer to sentences without claiming that what comes between the quote-marks is ever a sentence. So, quote-names do not refer to what is within the quote-marks, on this view. But it does say that they can refer to types of strings some of whose instances are sentence-tokens. This is no more surprising than the claim that the string-type "cat" can be tokened either as a noun-token, or, e.g., in the noun "cattle", where it is not a noun-token. In summary, if the compositionality principle is to be satisfied, we should say that in a token quote-name of a sentence, what comes inside the quote-marks is not a sentence-token. Therefore, (D) need no exception. (Those in doubt may of course read in an exception to (D).)

Another common objection against an unrestricted intersubstitutability claim is that whereas the truth-ascribing sentence containing a "that"-clause always commits itself to the existence of something which is true, the sentence in the "that"-clause need not (cf. Field (1994a: 250)). Thus, one might argue, in the sentence "That snow is white is true; therefore, something is true", one cannot replace "That snow is white is true" with "Snow is white". But this can be resisted. Suppose someone, inspired by Nietzsche, claims that *nothing is true*, and that a more analytically-minded philosopher tries to trap him by asking, innocently, "But is snow white?". He may then, if the Nietzschean says "Yes", go on to say, "But if snow is white, then something is true!". This is a fully reasonable response. More importantly, we do not take it to be assertible only on the assumption of other sentences. It is simply one that a competent speaker will assent to (absent semantically irrelevant inhibiting factors, like the occasional obstinacy of a Nietzschean). The explanation of this is that our understanding "true" simply consists in our speaking in conformity with (D).

But, the objection might continue, one cannot similarly substitute in the sentence "It is *logically* true that if that snow is white is true, then something is true". This charge fails because the definitions of logical truth typically make it a property of *sentences*, e.g., so that sentences are logically true iff true for all variations of the non-logical vocabulary, or true under all interpretations, or in all models, etc. But on such an explanation of "logically true", the sentence in question is strictly speaking unintelligible. Since "contains three words" is a property of sentences, the sentence "It is logically true that snow is white" is like "It contains three words that snow is white". Of course, we do, in philosophy, say things like "It is logically true that if *p*, then *p*", but if logical truth is a property of sentences, this must be taken a sloppy and misleading way of saying that "If *p*, then *p*" is logically true. But then, the above point about quote-contexts can be made anew, and the alleged counter-example fails.

Could one not define logical truth instead for *propositions* in such a way that a counter-example to (D) could be formulated on the lines above? Note first that, by (D), we easily obtain all instances of

The proposition that the proposition that *p* is true = the proposition that *p*,

by applying (D) to the left hand side of a self-identity sentence

The proposition that *p* = the proposition that *p*.

(By the same token, we easily derive the instances of "To say that it is true that *p* is just to say *p*"). Thus, (D) entails that there can be no definition of a property with the consequence that a proposition *P* may have (lack) it while the proposition that *P* is true does not. To get a counter-example to (D) one thus only needs to show that these pairs of propositions are distinct, but not necessarily by defining "logically true". Probably, opponents to (D) will mainly argue that they have different *structure*. I cannot here provide a rebuttal of this view, of course; suffice it to anticipate that the nominalist account of propositions to be proposed in Chapter 5 entails that the idea that propositions have structure is wrongheaded.

It is crucial that the equivalence expressed in (D) is a semantic notion, in the sense that it does not concern various pragmatic effects, poetic value, or connotation, that utterances of the relevant sentences may have. The distinction between pragmatic and semantic becomes especially relevant as

concerns the putative counter-examples pertaining to propositional attitude contexts like "*x* believes/knows/says/means that ...", "Sentence *s* means that ...", etc. To wit, we find the following to be possibly true:

X believes that snow is white but not that it is *true* that snow is white.

First, this intuition is apparently not shared by everyone, since Frege, for one, said, on the contrary, that always when one judges something to be thus or so, one judges a thought to be true (cf. 2.2). Unlike me, he could not be accused of adapting his intuitions to fit his theory, since he had no theory which entailed it, but simply found it intuitive. But many do have the intuition and this must be accounted for. The sentence above seems to say that *X* believes that snow is white, but not that "true" applies to this proposition, or some such. But this can plausibly be taken to be pragmatically, not semantically, expressed. What is semantically expressed is a contradiction, given (D). The pragmatic effect can then be taken to arise because of the following facts: Firstly, the semantic content is a contradiction, wherefore this interpretation is automatically avoided (hence, what is *saliently* expressed is *only* a pragmatic effect). That we should avoid interpreting utterances as communicating contradictions is common to all pragmatic theories. Secondly, the second conjunct contains a semantically redundant phrase, "it is true that", wherefore an interpretation is made which satisfies the assumption that the apparent violation of a maxim (under the category of *Manner* – "Be brief!") can be explained. Similarly, on the Relevance Theory of Sperber and Wilson (1986), it should be assumed that the utterance is the least effort-requiring way of communicating the message.

Therefore, the semantically redundant phrase will be taken to have a point. Thus, something more than what is semantically expressed will be taken to be communicated. Thirdly, since this phrase is semantically redundant, it must be either a non-semantic property of the phrase (like its poetic value) or (a part of) the very phrase itself, which guides the audience (us) to the right interpretation. Fourthly, the interpretation which both stays closest to the semantic content of the sentence and satisfies the assumption that there is a point with the semantically redundant phrase is something like the above interpretation, on which what is communicated is that *X* believes that snow is white but not that "true" applies to the proposition that snow is white (perhaps because *X* knows no English). Note that the interpretation is helped by the stress on "true". Without it, the intuition is less clear. But this is

also in line with modern pragmatics, according to which stress on a word indicates that it is to be taken to inform the interpretation in a way it would not do otherwise. Consider the sentence "*X* believes that blood is red, but not that it is true that blood is *red*". This is almost unintelligible, but might be interpreted as communicating (if anything) that *X* believes that blood is close to paradigmatic red but not *quite* paradigmatic red (or some such). The point is that it is the word "red" that guides the interpretation, rather than "true" in this case. In any case, I take it the above is a proper application of uncontroversial pragmatic principles that explain why the intersubstitution seems to fail in the relevant contexts.

4.3 THE VARIOUS OCCURRENCES OF "TRUE"

This section is devoted to arguing that (D), with a minor and similar-spirited extension, suffices for explaining the semantic contribution of "true" as it occurs in any context. Since (D) is taken to give the meaning of "true" entirely, the claim is that the semantic contribution of "true" is entirely given by (D). What will be done here is thus not to say how "true" contributes to the meaning of sentences containing it. Its contribution *is* its figuring in the equivalence described in (D). The point of this section, rather, is to show how one with (D) and other, independent facts about meaning and logic can explain how intuitively meaning-related facts about sentences with "true" can be explained. These facts consist mainly of intuitions about valid inferences and about "what has been said" by various sentences containing "true". We will thus not say that the "meaning" of, e.g., "Everything he said is true" somehow consists in, or is given by, its inferential relations to other sentences. Rather, these inferential facts are taken as consequences of the semantic property of "true", given by (D), and the semantic properties of the other expressions in the sentence. What follows, then, is *evidence* for the claim that (D) gives an exhaustive semantic characterization of "true".

Recall that the inferential constraint required that every argument depending for its validity on "true" be shown valid. That is, the conclusion should follow logically from the premises and the theory of truth. This will be shown for a number of such arguments, and these demonstrations are intended to make it plausible that any such argument can be explained on the basis of (D). To repeat, the explanations are meant to be neutral towards possible

explanations of the notion of inference (insofar as it is not explained in terms of truth, if this is indeed incompatible with deflationism). An important feature of this semantic account is that we do not give *paraphrases* or analyses lacking "true" of sentences containing it. There is no reason why this should be required, however, and we saw in 2.3 that imposing this requirement has caused much unnecessary confusion and disillusionment.

In the preceding section, occurrences in various problematic contexts like those created by propositions attitude verbs were treated on the basis of (D) and other independent (pragmatic) facts. The claim that (D) suffices for entirely capturing the semantics of "true" will now be further evidenced by treating six other types of occurrence, in this order:

(i)   in modified truth-ascriptions,

(ii)   in the "truth-operator",

(iii)   in quantified truth-ascriptions,

(iv)   in application of the truth-predicate to definite descriptions,

(v)   in demonstrative uses like "That is true", and

(vi)   in derivative uses like "true rebel", "true friend".

The next section is devoted to explaining what looks like truth-ascriptions to sentences. Except, perhaps, for case (vi), these explanations will more or less suggest themselves, in that it will be rather obvious which further facts about language to appeal to and how to exploit them in the explanations. What is important is *that* it is so obvious, and, therefore, how much more plausible (D) is than other deflationary theories, especially those that take sentences as primary truth-bearers.

It is crucial to the explanation of occurrences (ii)-(v) to note that "that"-clauses work in important respects like paradigmatic singular terms. In my view, they *are* singular terms, as will be argued at length in 5.4. In what follows, however, we need only assume that uncontroversial view that "that"-clauses behave in certain respects in a term-like way.

1. MODIFIED TRUTH-ASCRIPTIONS

Not many deflationists have addressed the question of how to deal with modified truth-ascriptions, such as "It might be true that snow is white", the exceptions being the prosententialists and Hartry Field. The formers' account, however, has been the target of severe criticism by Kirkham (1992: 327f.)), and for Field, the case of modified truth-ascriptions constitute the typical kind of impediment to the pure disquotationalist theory (as we will see in 4.4).

In showing how to treat modified truth-ascriptions, we may take intuitively valid inferences as our primary datum, for instance, that from "It might be true that snow is white" to "Snow might be white" and back. By "modification", I will also include such sentential operators as negation and adverbial modifiers like "necessarily", "presumably", "approximately", and so on. To see how the account will work, recall first (from 3.1) that (D) can be equivalently reformulated by quote-names of expressions and the Concatenation-function:

(D') Every sentence s is such that "that"^s^"is true" is S-equivalent to s.

Next, we introduce the notion of a *modifier-function* by these examples: the function $Neg(x)$ takes sentences to their negations; $Might(x)$, similarly, takes the sentence "Snow is white" to "Snow might be white", and so on; $Presum(x)$ takes "Snow is white" to "Snow is presumably white", and so on. Finally, we must also include functions whose values can only occur in wider contexts, such as $More(x)$, one of whose values is the first subsentence of "He is more swift than he is strong". The last kind is also necessary for dealing with subjunctive and counterfactual conditionals. We could now account for all these modified truth-ascriptions by the following extension of (D):

(MD) For every sentence s and modifier-function f, $f($"that"^s^"is true"$)$ is S-equivalent to $f(s)$.[2]

---

[2] Some adverbs, however, do not allow exportation from an ordinary sentence to the corresponding truth-claim. That is, you cannot infer, e.g., "That he ran is quickly true" from "He ran quickly", for the former does not make sense at all. It seems, then, that we must exclude those adverbs which, intuitively, are applied to the verb and not the whole sentence. In the case of "approximately", the step seems to be allowed in virtue of the

Now, a simple instance of (MD) is precisely that the sentences "Snow might be white" and "That snow is white might be true" are S-equivalent. Since the identity function is in the range of the quantifier, (D) is entailed by (MD). It is thus a stronger claim that (D), but a natural extension of it.

The claim (MD) actually has some wider implications, which deserve comment. Peter Smith (1998) has proposed a treatment of "is approximately true" on these lines, so as to show deflationism compatible with the idea in the philosophy of science, that scientific theories are only approximately true. *Mutatis mutandis* for propositions, to say that the proposition that Sicily is triangular is approximately true is to say that Sicily is approximately triangular, and so on for all other approximate truths. Could we also make sense of the idea that some theories are more true than others, thus explaining problematic aspects of the progress of science? To say that it is more true that Einstein was right than that Newton was, is, as follows from (MD), to say that Einstein was more right than Newton was, which seems intelligible. But we would also have to admit the intelligibility of sentences like "Space is more relative than absolute" and "Electrons exist more than the ether". If these are unintelligible, then, the Deflationist should simply say, so is the claim that the one proposition is truer than the other. A similar situation holds for numeric degrees of truth. If it is true to degree 0.1 that phlogiston exists, then phlogiston exists to degree 0.1.

It may be objected, however, that whereas we can make sense of the claim that Einstein's theory is more true than Newton's, it does not make sense to say that space is more relative than absolute. In any case, the claims seem different, and therefore, a counter-example to (MD). Of course, we must here focus on the claim that the proposition that space is relative is more true than the proposition that space is absolute. A theory can derivatively be said to be more true if it contains a greater ratio of true claims, which is not a problem for (MD). First, if space is relative and not absolute, then it should not be unreasonable to say that space is more relative than absolute. If this is what is meant, then (MD) is not refuted. Insofar as the implication of gradualness is to be made greater justice, however, I believe a deflationist must say that the claim that the proposition that space is relative is more true than the proposition that space is absolute, is strictly speaking as nonsensical as the claim

---

sensibleness of treating the adverb as somehow modifying the claim made by the sentence, rather than the activity described by the verb, as with "quickly". There must be some explanation to this difference between adverbial phrases even if Deflationism is false, and it therefore poses no problem for Deflationism specifically.

that space is more relative than absolute. The sensible gradualness cannot, strictly speaking, concern truth *simpliciter*, but something like truth-likeliness or truth-obviousness (i.e., how likely or obvious it is that something is true). It is telling that the best examples of plausible claims made by "more true" concern theories for which there is only very indirect evidence, or involve vagueness. In such examples, it is easier to "read in" something other than truth, like likeliness or obviousness, which may nevertheless be intimately and relevantly related to truth. In finding out which physical theory is true, we have only indirect evidence to go on, which may make a theory more or less likely to be true. Concerning vagueness, it is more or less clear that an object is in the term's extension, and so, more or less clear that the claim made by the sentence containing the vague term is true. We may here repeat part of the explanation to the alleged counter-example to substituting in propositional attitude contexts, that given the non-sensicalness of what is literally said, we add something to this content to make sense of the sentences involving "more true". This kind of "enrichment" is of course ubiquitous in language use.

For most cases of modified truth-ascriptions, however, (MD) seems correct. That the equivalence between truth-claims and denominalized sentences is so strong is in itself remarkable. The more modifiers one examines, modal and evaluative modifiers, intensional contexts, occurrences in counterfactuals and subjunctive conditionals, the more it seems that the general claim must be right, its audacity notwithstanding.[3] This is something that should make anyone hesitate to take sentences as primary truth-bearers, given that T-biconditionals must be supplemented with separate clauses for every modification.

2. THE (SO-CALLED) TRUTH-OPERATOR

Sentences prefixed by the expression "It is true that" can easily be explained by recourse to explanation of truth-ascriptions to "that"-clauses. It is a general fact that when a predicate or verb-phrase "φ:s", which is applicable without semantic anomaly to "that"-clauses (as opposed, e.g., to "is green") occurs in a sentence, "That p φ:s", then there is always a trivially equivalent sentence of the form "It φ:s that p". For example, "That p is good" and "It is good that p" have equivalent instances, and the same holds also for more

---

[3] This transparency of truth-ascriptions has also been noted by Brandom (2000: 162f.).

complex expressions, e.g., "That *p* was not discovered until 1830" and "It was not discovered until 1830 that *p*". The best way to explain this is to regard the "It" in the latter type of sentence as anaphorically (actually, *cata*phorically) referring to the "that"-clause. These sentences are thus grammatically on a par with "He is nice your husband", which, of course, is equivalent to "Your husband is nice".[4] As Horwich (1998a: 16, n. 1) notes, the grammaticality of the sentence "It is true what Oscar said" is further evidence for this view. Thus, (D) stands in no need of further addition. It should now be evident why expressions of the form "It φ:s that" should not be called *operators*: it suggests that a sentence "It φ:s that *p*" is primarily formed by "It φ:s that" and the sentence following it. This analysis clashes with the strong evidence that "that"-clauses are singular terms, and so separate syntactic units (cf. 5.4).

3. QUANTIFIED TRUTH-ASCRIPTIONS

To see how to explain quantified truth-ascriptions, consider the argument we looked at in 1.5:

(B)   Nothing Descartes believed is true.
       Descartes believed that he existed.
       ──────────────────

       Therefore, Descartes did not exist.

---

[4] McGrath (2003: 668, n. 9) notes that many sentences embedding "that *p* is true" are awkward, and that we therefore often prefer "it is true that *p*". For instance, "He believes that it is true that *p*" is clearly to be preferred over "He believes that that *p* is true". There can be no question that "that *p* is true" is grammatical, however, and so it must also be correct to embed such a sentence in sentential contexts. I believe some of these embeddings seem ungrammatical because they have "centre-branching" trees (cf. Yngve (1960)). Such sentences are relevant for linguists because they seem clearly dictated to be grammatical but nevertheless appear unintelligible. This is taken to be because of the amount, or kind, of cognitive processing that is required to parse them. An example is "The man the boy the girl kissed met ran", which "should" just mean approximately, "The girl kissed the boy who met a man who ran.". As can easily be tested, the sentence-form "*x* believes that it is true that snow is white" is less centre-branching even on the grammatical analysis which suits this explanation worst (within reasonable bounds). This is plausibly the explanation to why such sentences are preferred to the former type, and why the former type appears awkward.

To explain this, we only need to look at an argument without "true" but of the same form: "Nothing in the room is mine, the dictionary is in the room; therefore the dictionary is not mine". Similarly, we can infer from the two premises in (B) "That Descartes existed is not true", and, then, by (D), "Descartes did not exist". This treatment explains how we derive, from sentences containing "true", sentences lacking them, as in the examples that showed the truth-predicate to increase the expressive power of a language (thesis VI, 1.4). Such strengthening is exemplified by "Nothing Descartes believed is true" in that it implies all sentences of the form "If Descartes believed that p, then not-p".

4. APPLICATION OF "TRUE" TO DEFINITE DESCRIPTIONS

Another kind of truth-ascription considered as an example of the fact that "true" increased the expressive power of English was "What Percy said is true". As with quantified truth-ascriptions, this will be dealt with by comparison to a sentence containing neither reference to propositions or truth, e.g., "What he bought is expensive". Now, just as the argument "What he bought is expensive, what he bought is (=) the last umbrella; therefore, the last umbrella is expensive" is valid, so is this one: "What Percy said is true, what Percy said is that snow is white; therefore, that snow is white is true". Both of these argument have the following form: $F(a)$ and $a = b$; therefore, $F(b)$. From the conclusion of the latter, we get "Snow is white", and this shows why the original truth-ascription has all instances of "If what Percy said is that p, then p" as consequences (Horwich (1998: 3)).

5. APPLICATION OF "TRUE" TO DEMONSTRATIVES

The simple phrase "That is true", as uttered in response to someone's utterance of a sentence is plausibly explained in a similar way. In the previous example, one might say that, given a sentence identifying the thing said (or, for other cases, believed, assumed, feared, etc.), that is, a sentence of the form "What x said is that p", we can, with the truth-ascription, infer a corresponding denominalised sentence. When we say in response to some utterance, "That is true", this should be taken as short-hand for "What has been said is true". Given that it is normally obvious what has been said, there is a sentence of the form "What has been said is that p", such that both speaker and audience can infer, together with the truth-ascription, the sen-

tence "*p*". It is probably because it is normally evident to both speaker and audience which denominalized sentence to infer from "That is true", that some philosophers have explained utterances of "That is true" as simply being equivalent to the foregoing sentence (especially Strawson and the prosententialists – cf. 2.5 and 2.8). But this obscures the fact that two steps have been taken and, crucially, it does not explain this equivalence from a unified theory on which other truth-ascriptions can be explained.

When the utterance preceding this type of truth-ascription contains context-sensitive expressions, the sentence the audience will infer from the truth-ascription will not always be identical to that uttered. If the foregoing utterance is, e.g., "I am hungry", I will not infer "I am hungry" from my utterance, "That is true". But this phenomenon is general. If you utter "I am hungry", and I respond, "I believe that, too", it is incorrect to say "I believe that I am hungry" is an appropriate inference from these utterances. This is naturally taken to be explained by reference to the propositions expressed, of course, but on the theory of propositions to be sketched in the next chapter, such explanatory appeal to propositions is forbidden. We therefore have to assume that speakers accomplish something like a tacit translation of context-sensitive expressions when they occur in utterances other than their own present ones, into sentences that, intuitively, say the same thing as uttered in the present context. Whether the view of propositions is correct or not, the phenomenon is general and not produced by the truth-theory, and so we may assume there to be a solution to it, which, together with (D), will entail that from my utterance "That is true", as a response to your utterance "I am hungry", one can infer "You are hungry", as uttered by me in the same context. What has to be done here is to take the notion of inference to hold relative to contexts in such a way that the sentence "I are hungry" can be inferred from "You are hungry" relative to a context in which the speaker uttering the first is the same as the one to whom the second is uttered. That is, for some sentence-pairs, the one can be inferred from the other only relative some contexts. For the above pair, this will hold only in contexts where the personal pronouns are assigned the same values. With this notion of inference, (D) can be taken as it stands and yield the desired result: from "That is true", as uttered in response to "I am hungry", "You are hungry" can be inferred.

6. DERIVATIVE APPLICATIONS OF "TRUE"

Finally, we will see that (D) even enables a promising account of such derivative uses of "true" as in "He is a true rebel". First, what are the intuitions about this sentence that an explanation should do justice? It seems reasonable to take this to be semantically equivalent to, "He is a rebel". For one thing, they intuitively have the same truth-conditions. It would not be true to say that he is a rebel, unless he were a true rebel. However, the former clearly implies something more. To me, this use of "true" seems to be used to refer to things that *do not merely claim to be such and such*, *but really are*.

To see how to account for the semantic intuition, we can treat the former sentence as on a par with "He is an obvious criminal". This cannot be analyzed as "He is a Danish prince", because what is said is not that he is obvious and criminal. Rather, it is elliptic for "That he is a criminal is obvious". Likewise, "He is a true rebel" does not say that he is true and a rebel, but should be seen as elliptic for "That he is a rebel is true". The pragmatic difference is due, in the first instance, to the semantic redundancy of "true", which violates conversational maxims (cf. the argument in 4.2 on applying (D) to propositional attitude contexts). As suggested to me by Peter Pagin, the pragmatic effect of "true x" might be explained by contrasting it with "false x". Now, "false x" is not used to say simply that something or other is not an x, but to say in addition that the object is falsely indicated or said to be an x. A false coin is falsely indicated to be a coin, and a false friend is falsely indicated (by himself) to be a friend. The question is then how this effect is produced. If this can be explained, then "true x" may simply be explained as communicating the negation of what is communicated with "false x". What is negated is plausibly a conjunction: that the thing is not an x and that it is indicated or said to be an x". This, together with the semantic content, squares nicely with the intuition that the expression "true x" is used to say that the thing is not merely indicated or said to be an x, but really is.

Let us now take falsity simply to be lack of truth (deflationism is well-known to leave few options here). Now, we can see that the expression "false x" does two things. Firstly, as noted above, a sentence "y is a false x" is elliptic for "That y is an x is false". On the semantic side, therefore, "false x" contributes to communicating that y is not an x, by (D). Secondly, on the pragmatic side, it contributes to communicating that the object is said or indicated to be an x (which, by the semantic content, is false). We may next note that when propositional properties are superficially ascribed to non-

propositions, the interpretation is that it is something said or indicated which has the property. When we say, "I believe *her*", it is, non-elliptically speaking, *what she said or indicated* that is said to be believed. A person cannot, strictly speaking, be believed. Since, in "*y* is a false friend", a person is ascribed the propositional property of falsity, the interpretation is that what the person says or indicates is false. What is it, that is said or indicated? Since the sentence "*y* is a false friend" is, in the first instance, an ellipsis for "That *y* is a friend is false", the proposition in question is plausibly precisely that *y* is a friend. And so it is that the conjunction mentioned above is communicated.

A word of caution is that this use of "true" may have become lexically distinct, at least for some speakers, though once originated in the pragmatic way here envisaged. This is of course the case with many expressions, e.g., dead metaphors. In that case, (D) would not entirely exhaust the semantic meaning of "true", though it would figure essentially in the causal-historical explanation of the additional sense.

4.4 THE TRUTH OF SENTENCES AND UTTERANCES

We have seen that deflationary theories taking sentences as primary truth-bearers fails in accounting for ordinary intuitions and use. The approach taking "true" to be applied primarily to "that"-clauses, by contrast, need only be supplemented with more or less truistic claims about language in order to account for sentences containing it. One type of occurrence that will naturally attract special attention, is where truth seems to be ascribed to sentences. I take these occurrences to be elliptic variants of the type of truth-ascription exemplified by "What Percy said is true". To wit, Sentence-Truth is to be explained on the basis of,

(ST)   "(The sentence) *s* is true" is elliptic for "What *s* says is true".

This may be contrasted with a different deflationary account of sentence-truth,

(ST') If *s* means [i.e., says] that *p*, then *s* is true iff *p*,

which has been discussed, e.g., by Soames (1999: 238). Now, (ST') is clearly a consequence of (ST), but, since it only gives a sufficient condition for a sentence to have certain truth-conditions, it is not obvious how to explain various truth-ascriptions to sentences on its basis. It therefore seems plausible to agree with (ST') but take it to be explained by (ST).

It is plausible that all so-called "derivative" uses of a predicate, i.e., a use for something that is not of the right category for the word, is elliptic. Thus, one says that one has a red visual experience, this is elliptic for "One has a visual experience of something red", or some such. Closer to (TS) is the phrase "believe a person", which is elliptic for "believe what a person says". As noted in 2.11, saying that sentences are "derivatively true", etc., is misleading because it implies that sentences are true, though in a special way. But if something else than sentences are taken as primary truth-bearers, then sentences should not, strictly speaking, be taken as true at all, just as experiences should not in any way be taken to be red. Ellipses like these work because of pragmatic inferences: since it is absurd to say literally that an experience is red, the audience finds a pragmatically sound interpretation as close as possible to what is literally said.

It seems plausible that there is only one primary truth-bearer. If not, then we have two different truth-concepts, which seems gratuitous, since truth for different entities are more or less straightforwardly interdefinable. However, defining sentence-truth in terms of propositional ditto is remarkably simpler than the converse, which is of course one of the main arguments for the present standpoint. It is not reasonably considered unimportant what we take as primary truth-bearers, given how central truth and meaning are to analytic philosophy. Indeed, this choice seems pivotal in that taking propositional truth to be primary seems to exclude a truth-conditional theory of meaning. If propositions are primary, then what it is for a sentence to have certain truth-conditions (in a context) is for the proposition it expresses (in that context) to be true under those conditions. But one way of saying what a theory of meaning is, is by saying that it is to explain what it is for a sentence to express a proposition (relative to a context). So, in order to explain what it is for a sentence to be true (and thus have certain truth-conditions), we must explain what it is for it to express the proposition it expresses (relative to a context). The concept in terms of which meaning was to be explained thus involves the concept to be explained in an immediate way, and so the theory moves in a narrow circle. Taking propositional truth to be primary therefore excludes explaining sentence-meaning in terms of truth-conditions (cf.

Dummett (2003: 5) and Soames (1999: 244), whose reflections on these matters come very close to this conclusion).

The expression "say" in (TS), is plausibly also taken to be derivative. Sentences do not say anything, people say things in uttering them. Furthermore, what people say in uttering a sentence depends on the context. But precisely because what is said by a sentence depends on the context, truth itself need not be relativized to contexts, if we take propositions as primary truth-bearers. On such an account, the context-dependence of sentence-truth derives from the context-dependence of saying. Theories taking truth to be a property of sentences, by contrast, must take truth itself to be relative to contexts. Taking propositions or utterances to be primary therefore has the advantage over primary sentence-truth of leaving truth absolute. (True, not all relativizations of truth are equally counter-intuitive; so-called "temporalists" intuit that the truth of propositions is relative to times. My point is that sentence-truth entails certain counter-intuitive types of relativity.)

Truth for sentences must also be relativized to languages (the awkwardness of which was urged by Black (1948)). Clearly, a sentence may be true in one language and false in another. This phenomenon is of course also obvious for the propositionalist: sentences express different propositions in different languages. Propositional truth should also be favoured over *utterance*-truth in that truth is not plausibly taken to be a property of an event (many, of course, also think it is confused to say that a string of letters or phonemes, i.e., a sentence, can be primarily true). Further, defining propositional truth in terms of utterance-truth is unnatural and requires non-actual utterances, whereas the converse is simple and obvious. It is also hard to see what a theory of utterance-truth would look like. One is rather tempted to say that an utterance is true iff something true is said by the utterance. Indeed, in some cases, those favouring utterance-truth have just conflated the thing said by an utterance with the utterance itself. It is clear that Austin (1950: 113f.), Strawson (1952: 2ff.) and other similar-minded philosophers at the time were analogously unclear about the difference between the act of stating something and the thing stated, both properly called "statements", a conflation which considerably weakened their case against Tarski's sententialism.

We will now look at a few further cases of truth-ascriptions to sentences which meant trouble for disquotationalism, but which are easily treated on the present account. The cases of context-sensitivity is obvious: "I am hungry" expresses different propositions depending on its speaker, and

therefore has different, corresponding truth-conditions. Further, certain types of modal and counterfactual truth-ascriptions have been discussed at length in relation to disquotationalism[5], in particular,

(M1) If we had used the word "white" differently, "Grass is white" might have been true,

(M2) If we had used the word "white" differently, grass might have been white,

(M3) "1+1=2" might not have been true.

On Field's original, "pure disquotationalism", (M1) was taken to be equivalent to (M2), though, as we saw in 2.11?, he later proposed a truth-theory quantifying over meanings to give a more intuitive account. With (D), (MD) and (TS), we can avoid the dilemma by noting that there are two ways of spelling out the ellipsis referred to in (TS). Depending on the reading, the Consequent of (M1) is one of the following:

(C1) What "Grass is white" (actually) says might have been true,

(C1') What "Grass is white" would (then) have said might not have been true.

Assuming some obvious facts about what "Grass is white" actually says, we see that, on the first reading, (M1) entails (M2), but is not true, and on the second, it is true, but does not entail (M2). Concerning (M3), Field notes that there are two readings here, one on which, intuitively, it says that what "1+1=2" (actually) says might not have been true (absurd), and one on which it says that what "1+1=2" might have said (in certain circumstances) is not true (plausible). So, whereas simple truth-ascriptions to sentences are given by (TS), modified ones allow readings with different scopes of modal and subjunctive expressions. It is of course the former reading of (M3), and (C1) which Field was after when spelling his truth-predicate "true-as-we-under-stand-it" (1994a: 265f.). This intuitive way of specifying the intended reading

---

[5] See, e.g., Etchemendy (1988: 63f.), Field (1986: 58) and (1994a: §§ 5, 9), García-Carpintero (1998: 50ff.), Halbach (2001) and (2002).

is of course easy to explain: we understand it as *saying that grass is white*. With this in mind, the derivation of the disquoted sentence is unequivocally valid, since the other interpretation is explicitly ruled out.

We will end this section by discussing how (ST) fares with sentences that do not intuitively say anything, or say more than one thing. We should first note that ellipsis is a pragmatic phenomenon, and therefore does not obey strict rules that are to apply in every case. However, (ST) does seem to work also for most problematic cases. Take, to begin with, a sentence which does not seem to say anything because of non-existing words, such as "Gweeks thwartle" (assuming this to be a sentence at all). One may have the idea that the claim that this sentence is not true is itself true. Since nothing is said, there is nothing than can be true. This seems to me to be primarily a philosopher's reaction, however. A normal speaker would find this merely confusing. One might be prepared to say that "Snow is black" is indeed not true, but since "Gweeks thwartle" is meaningless, it is simply confusing to say that it is not true. One may of course change one's mind if it is explained that *since* it is meaningless, it cannot very well be true. But such a claim would enforce a different interpretation, namely the interpretation on which "*s* is not true" is interpreted as saying that *s* does not say something true, rather than (by (ST)) saying that what *s* says is not true. That such opposing reactions are manifested is rather natural given that it takes pragmatic interpretation to make sense of ascriptions of sentence-truth in the first place.

(ST) registers a default type of interpretation. This is why a speaker will at first experience as awkward the claim that "Gweeks thwartle" is not true. If (ST) is applied here, then, since nothing is said by the sentence, the claim will be taken as a supposition failure.[6] When a string of letters is said to be true or not true, it is presupposed that it says something. The opposing interpretation, which is intended when it is explained that the sentence is

---

[6] An adherent of Russell's analysis of definite descriptions would say that the confusion in this case is caused by a scope ambiguity of the negation. I will not here discuss this disagreement at length here, since the differences for this case are only special cases of the general, well-known differences between these competing analyses of definite descriptions. In my view, Russell's analysis gives the wrong predictions for sentences with empty descriptions and an external negation, e.g., "It is not the case that the kind of France is bald". On his analysis, this should be unproblematically true, since there is no possibility of scope ambiguity for the negation. But this sentence appears just as "inappropriate" or "failed" as "The king of France is bald". In any case, I will assume the presupposition view to be correct.

untrue since it does not say anything, is special, since it is usually trivial and irrelevant that meaningless sentences do not say something true (since they do not say anything at all). As always, such pragmatic considerations of relevance govern interpretation, whence the possible confusion upon encountering the claim that "Gweeks thwartle" is not true. If a philosopher for some reason wants to say that meaningless sentences are untrue, he is entirely free to do so. He will then just make clear that when calling a sentence true, he means to be saying that it expresses a true proposition. The claim (ST) is not legislative, but records a fact about normal speakers' interpretation, which also tends to be philosophers' interpretation, absent various theoretical involvements that enforce a different interpretation. We have already seen that though (ST) works for most normal cases, it does not determine the interpretation for any possible case, as a *semantic* principle should. The modal truth-ascriptions were a case in point. The deviating cases are explained by reference to ordinary pragmatic principles.

Sentences which do not say only one thing are treated similarly. With cases of ambiguity, as in "Banks are man-made", truth-ascriptions are judged relative to the interpretation of "bank". If a particular interpretation is salient, (ST) might be directly applied. Then, what is said is that what the sentence says under that interpretation of "bank" is true (or not). If no particular interpretation is salient, the truth-ascription will raise the kind of confusion as did "Gweeks thwartle". Is the truth-ascription taken to be a supposition failure, or is it to be taken as true, since on one interpretation of "bank", the sentence says something true, or is it false, since on one interpretation of "bank, what is said is false? What is meant in a real situation, if that can be worked out, is of course determined by contextual features, but taken in isolation, there is no real question here, since there is no determinate interpretation. That we normally interpret truth-ascriptions in accordance with (ST) is evidenced, however, by the fact that we tend to ask for a specification, rather than answering "Yes" or "No". If I ask "Is the sentence 'I am tall' true?", the audience will either presuppose a reference for "I", e.g., myself, and answer "Yes" or "No", or the audience will object, "It depends on who utters it". It is improbable that a speaker says "Yes" for the reason that *something* that could be said by the sentence (namely if uttered by a tall person) would be true. Thus, (ST) works most of the time, but because this is a pragmatic issue, we need not, and should not, say anything more specific. As long as there is some pragmatic explanation of every piece of linguistic

behaviour with respect to truth-ascriptions to sentences which takes speakers to interpret "true" as primarily ascribed to propositions, we are in the clear.

CHAPTER FIVE:
THE SEMANTICS OF "THAT"-CLAUSES

5.1 INTRODUCTION

It is often because of nominalist scruples that philosophers have spoken of truth for sentences rather than propositions, despite the counter-intuitive consequences noted in 4.4. For a deflationist, the choice is quite different, since deflationism coupled with the view that sentences are primarily true – that is, disquotationalism – fares especially bad in accounting for our ordinary use of "true" and is in general quite counter-intuitive, as we saw in 2.11. I am myself sympathetic to a nominalist stance, why I take comfort in the fact that (D) is not committed to propositions. It only claims an equivalence relation between certain forms of sentences. Since "that"-clauses uncontroversially exist and must have some semantic functioning to be accounted for, nominalists should have no ontological qualms about (D).

The aim of this chapter is thus to argue that the deflationary theory defended in the last chapter can plausibly be combined with a nominalist view of "that"-clauses, thus solving problem (3) of 4.1, i.e., the problem of "steering between the Scylla of admitting propositions and the Charybdis of disquotationalism". Of course, this requires that one give a linguistic kind of deflationary theory, like (D), but this was found mandatory in Chapter 3 in any case. Given the argument of Chapter 3, then, nominalist deflationists are free to adopt any view of the primary bearers of truth. Given the many advantages of deflationist propositionalism, and given that "that"-clauses are there to be accounted for in any case, the choice should be obvious. Some space will be devoted to the formulation of a nominalist constraint, which is to guide us in making sense of ordinary speech (and those of its features that engender realist metaphysics) without admitting propositions through the back-door. In

Berkley's words, we will show how to "think with the learned, and speak with the vulgar".

I will first provide a clearer statement of the nominalist constraint I will adopt, and defend this particular formulation of nominalism. Then, I will argue that the views of truth and semantics advertised in this work are in better shape for construing an account of "that"-clauses meeting the nominalist constraint than does ordinary, truth-theoretic semantics. It will also be seen that the conclusion of Chapter 3 about how a deflationary theory must be formulated is not happily combined with a non-nominalist view of propositions, which is thus a converse case of motivation. This mutual support is taken to display an attractive coherence of the deflationist-nominalist view. We will then discuss various more specific syntactic and semantic questions about "that"-clauses, such as the question whether they are singular terms, and their role in propositional attitude ascriptions and quantifications. As a corollary, a solution to problem (4) of 4.1 will be presented, i.e., Gupta's problem of explaining general facts about truth. It will also be seen that this avoids a problem that arises for Horwich. After that, we will briefly discuss the term "proposition" and its relation to "that"-clauses. Finally, I will say something about what I think a correct semantics should be like, given the deflationist and nominalist theories presented. The main contribution this chapter makes to the preceding discussion, however, is showing how to get the good from propositional deflationism while avoiding commitment to propositions, the latter being the main motivation for disquotationalism.

5.2 THE NOMINALIST CONSTRAINT

In view of what was said in the previous chapter, that one must take propositions as the primary truth-bearers, it may be surprising that a nominalist view of propositions will now be defended. Admittedly, I have spoken "of propositions" above in the sense that I have been using the term "proposition" in an obviously committive way, i.e., as entailing by existential generalization that there are propositions. The nominalist theory that will here be sketched, however, takes such commitment to be forgivable, or even mandatory, in casual speech and in many fields of philosophy, but holds that this usage is to be treated as a datum that should be explained in a nominalist way, rather than something that must be eliminated. Some might not want to call such a view

"nominalist", but I take it to be sufficient for deserving the label that it is motivated by the common nominalist grounds, and that it is in opposition to non-nominalist theories (except, possibly, to such "two-dimensional" views of ontology as Carnap's (1956)).

At one level, the type of ontological commitments made in the previous chapter must be done away with. A beginning of such a regimentation was the alternative phrasing of the view that it truth is primarily a property of propositions, namely, *"true" applies primarily to "that"-clauses*. Rephrasing the claims and arguments of the previous chapter should not seem anymore worrying than the general requirement to give a nominalist treatment of all uses of "that"-clauses and what appears superficially to be quantification over, and reference to, propositions, which is ubiquitous in natural language. (The word "proposition" is not used in natural English as it is in philosophy, of course, but well expressions like "things we believe, say, etc.".) We may say that my above reasoning using "proposition", etc., is an instance of *semantics in the material mode*, which is to be regarded as a preliminary to the final, nominalistically kosher semantics. It is a convenient way of displaying linguistic intuitions and inferential relations by using, rather than mentioning, various sentences. The final semantics is to explain the facts about our linguistic use so displayed, but may *not* use the committive expressions, but only mention them. The reason that no "that"-clause may be used is that a sentence containing it entails a corresponding existential sentence. For instance, "I believe that snow is white" entails "I believe something". The word "proposition" is simply used (by philosophers) in such a way that a sentence like "I believe something" is a direct commitment to propositions, since propositions just are the things believed, known, etc. (More on the philosophical use of this word in 5.6.) We may now more explicitly formulate the Nominalist Constraint:

> (NC) There must be no quantifying over, or referring to, propositions and no use of notions primarily defined for propositions.

The nominalist claim that I will here argue for is that we can give an exhaustive account of all uncontroversial facts while still obeying (NC). By "uncontroversial fact", I mean to exclude such alleged facts as that there really are things we believe, assert, etc., with some nature to describe. It is of course question-begging to require a nominalist to explain this alleged fact. The "uncontroversial facts" include, rather, those about our linguistic behav-

iour and cognition in general that can be expressed in a nominalistically tolerable way. Particularly relevant for the present study, then is the question whether the semantics for "that"-clauses, quantifications like "Everything he believes...", and words that apply primarily to "that"-clauses, like "true" and "believes", can be given in a metalanguage which obeys (NC).

By disallowing "quantifying over" propositions in (NC), I mean that no quantified sentence whose instances contain "that"-clauses may be used in the semantics. For instance, "He believes everything he knows" has instances like "If he knows *that snow is white*, then he believes *that snow is white*". It seems appropriate to explain the commonly used "quantify over *x*'s" in this way, since the latter seems to entail that there are things, namely propositions, that we must not quantify over. This would make the constraint violate itself. Reference to propositions, similarly, is to be understood simply as the *use* of a "that"-clause, or a description like "What he said", which can intelligibly be inserted together with a "that"-clause in an identity-sentence ("What he said is (=) that snow is white"). Finally, it is only natural that a nominalistically acceptable semantics not include in its conceptual equipment (or "ideology", in Quine's sense) any "notions defined primarily for propositions", that is, words that apply primarily to "that"-clauses. The "primacy" here could be explained in terms of learning the word or introducing it in the language. If a word can only be learnt or introduced by figuring in sentences that also contain a "that"-clause, then it is primarily applied to "that"-clauses, on this terminology. According to the above deflationary theory, where (D) plays an essential role, "true" clearly applies primarily to "that"-clauses in the relevant sense, and it is rather obvious that propositional attitude verbs do as well.

This type of nominalism should be distinguished from what may be called *traditional nominalism*, the simple claim that there are no propositions. I have avoided this formulation because I think it misleading in important ways. In my view, this claim may be regarded from two different perspectives (not wholly unrelated to Carnap's distinction between internal and external questions (1956: suppl. A)). As expressed by (NC), we may not, when constructing the semantic theory, use "that"-clauses, quantified sentences with "that"-clauses as instance-terms, or expressions which are primarily applied to "that"-clauses (like "true"). This plausibly means that the "proposition" will not be in the semantics either. The sentence "There *are* propositions" will thus be compatible with the semantic theory, but I believe this is less of a problem and that adhering to (NC) is anyway a recognizably nominalistic standpoint. The reason for this roundabout formulation of nominalism has to do with the

second "perspective" from which the denial of propositions may be taken. This is related to what I called "semantics in the material mode". This is the activity of uttering those sentences that are assertible and drawing the correct inferences in order to see which intuitions a final semantics should account for. Otherwise put, one hereby displays facts about meaning that the semantics is to explain. It is simply much more convenient, and, more importantly, a safer way to avoid confusion, thus getting unreliable data, than by discussing speakers' use in a more theoretical (and nominalistically acceptable) way. For instance, asking a speaker, "Does someone who believes that snow is white always believe something?" will give clearer results than asking him to answer metalinguistic questions about the relevant sentences. The semantics, of whatever kind, should simply explain ordinary linguistic use, not high-level philosophical discussions about semantics. It was mainly by the "method" of semantics in the material mode that we in the previous chapter found that it is more convenient to take "true" to be primarily applied to "that"-clauses, rather than quote-names of sentences. But this result was reached mainly by asking whether propositions were more plausibly taken as the primary truth-bearers than sentences, not exclusively by asking metalinguistic questions.

When doing semantics in the material mode, it is of course necessary to assent to sentences with "that"-clauses and propositional quantifications. We are here simply to act as normal speakers, these being (by definition) the object of the semantic study. The semanticist is included in the subject-class, and, in practice, typically the only subject. Hence, it is clearly incorrect to assert traditional nominalism if we are doing semantics in the material mode. In fact, if we use "true" in accordance with (D), as I have argued is what underlies every use of the word, then any sentence "p" entails the corresponding "That p is true". But then, given that "that"-clauses admit of existential generalisation, any sentence whatsoever entails that there are propositions. Any sentence thus contradicts traditional nominalism. This does not mean that we have to admit propositions after all, only that we should be careful with the claim that there are no propositions. The analogy with Carnap's distinction is obvious, but I would like to emphasise that the two "perspectives" that I appeal to do not correspond to alleged distinct interpretations of "there is" or "exist". This alleged ambiguity is not empirically supported and does not have intuitive support, but is suspiciously handy for eating the ontological cake and having it.

Another obvious reason that care must be taken with this claim is that when doing other kinds of philosophy, it will be practically inappropriate to follow (NC). For instance, when doing epistemology, these ontological questions are irrelevant and should therefore not encroach by requiring cumbersome reformulations. For instance, instead of saying that one believes what one knows, (NC) would have us say that any sentence "*x* knows that *p*" (analytically) entails the corresponding "*x* believes that *p*" (or some such), but no nominalist will be able to follow the requirements on her semantic theory in this way. Thirdly, of course, no one will be able to obey (NC) in ordinary life. Therefore, we might as well anticipate possible criticisms by separating what is acceptable in a nominalist semantics from what may be accepted in more informal circumstances. One might say that the above formulation of nominalism is intended (partly) to avoid typical charges of self-defeat.

However consistent, this may appear as some kind of fiddling. Some will probably exclaim, "But are there propositions or not?". My instant ascent to a metadiscursive level may seem dishonest or evasive, but, in my opinion, it is really only a move necessitated by the fact that every philosopher and nominalist is also a normal speaker, who is not always, not even always when philosophizing, doing semantics. There is no theoretical problem with the kind of wide "inconsistency" allowed here – that is, it is not an "inconsistency" that indicates that the nominalistic view is false or unworkable, i.e., that we need to posit propositions in order to account for the uncontroversial facts.

Should we, *when doing semantics*, say that there are no propositions? Well, it is ruled out by (NC), since any use of a word introduced together with a "that"-clause is disallowed. And I think this is correct. When doing nominalist semantics, one should act as an observer and commentator of what has been called the uncontroversial facts, among which is the fact that speakers utter "that"-clauses and "things said/believed/known", etc., in certain conditions, and try to find a systematic explanation of this behaviour. He will thus mention all the words that are ruled out by (NC), and explain our use thereof, but it is question-begging to object that he must *use* these expressions in the explanations. That he need not is precisely the nominalist claim. If one engages in conversation with a speaker in order to display fact about usage, or, as is more common, if one tests what intuitions one has oneself, one must temporarily abandon the role as semanticist. As long as the (somewhat imaginary) orthodox nominalist can be seen not to end up in contradictions

himself, it should not be unproblematic to allow him to use the expressions in order to "experiment" with other speakers or himself.

It may seem as if a traditional nominalist must say that we ordinary speakers somehow go wrong every time we use a sentence with a "that"-clause. But this would surely go counter to a plausible principle of common sense. Like Cartesian scepticism, it seems to set the standards too high. This is not the type of argument that will persuade a real sceptic, of course, but there are very few sceptics, and precisely because of its clash with common sense. The issue concerns not how to persuade someone with radical views but the fact that most of us do accept the constraint on philosophical theories that they accord with common sense. Traditional nominalists therefore usually do try to account for ordinary judgments, by taking the commitment to propositions to be only an illusion created by the superficial form of the relevant sentences. The real meaning of the relevant sentences is then thought to be given by various paraphrases which lack commitments, and the view is often called "paraphrase nominalism". These analyses have been much criticized, however, and widely abandoned (below, I will myself explain why I do not accept them). If paraphrase nominalism fails, the clash with common sense is another reason for adopting the more "roundabout" nominalism.

But if we adopt (NC), do we not still somehow have to say that people are massively wrong? One idea may be that we have to say that people all say false things when uttering "that"-clauses. But falsity, like truth, is a property of propositions, and is therefore not to be used to characterize ordinary language at all in the nominalist semantics. Any sound semantics should make people come out mostly right, but "right" must not be interpreted as "true", if (NC) is to be respected. According to the semantics that will be defended (and which is in any case mandated by deflationism), uttering "I believe that snow is white" is correct in precisely the same sense as it is correct to utter "Snow is white". They only have different types of correctness conditions: when stating the correctness conditions for the latter we may (and should, I believe) use "snow", but in stating those of the former, we may not use "the proposition that snow is white".

Even if the present nominalism does not clash so obviously with common sense as does the traditional variety, there may still linger an initial doubt that such a linguistic take on the phenomenon of proposition can be right. Here, we could repeat the line of reasoning in 3.4, and appeal to those issues where only a linguistic answer seems possible, i.e., concerning "what it is to exist", "the average person" and "sakes". Even given a full account of the existential

quantifier and the existence-predicate, someone may ask, "But what is it to exist? You have only described the related linguistic expressions!". This type of charge should not without further argument seem more pressing for the present proposal that it does for a "linguistic view of existence or the average person". That is, we should require further argument for the claim that the notion of proposition in particular requires more than a merely linguistic account. It is not obvious an argument to that effect, which does not beg the question against nominalism, can be construed.

5.3 THE COHERENCE OF A NOMINALIST-DEFLATIONIST VIEW

In order to do justice to ordinary linguistic intuitions and use, a positive semantic value must be ascribed to all or most sentences that we unconditionally assent to. In a non-alethic semantics, one can take this value to be the very assent itself, or one may take it to be some notion of (unconditional) assertibility or analyticity (cashed out in non-alethic terms). An ordinary truth-theoretic semantics will typically take it to be truth proper. We may for simplicity disregard such notions as supervaluationist truth, high degree of truth and other such variations. First, they seem to get similar results for the cases that will be discussed, at least if these truth-notions are even remotely similar to ordinary truth. Secondly, they are dialectically rather peripheral and have cogently been objected to on more principal grounds.

We will first see that a truth-theoretic semantics that satisfies (NC), while accounting for uncontroversial facts about usage, will have rather implausible consequences, and therefore seems unsuitable for being coupled with nominalism. It will next be argued that a use-theoretic type of semantics does not seem to be in danger in these respects. There might be other reasons yet to be given why such a semantics cannot be both adequate and satisfy (NC), but, *prima facie*, it is particularly truth-theoretic semantics that seems unfit for nominalism. It will emerge after this discussion that the preferred alternative semantics will in important respects play a polemic role, in that its potential success in accounting for realist's intuitions is simultaneously a weakening of the realist's case. Finally, we will see that given the way a deflationary theory must be formulated, as concluded in Chapter 3, there is a reason why no deflationist ought to admit propositions. So, while the first consideration

shows why nominalists should love deflationism (or at least the kind of semantics it requires), the second shows why deflationists should love nominalism (whether they like it or not). There is thus a mutual coherence in the combination of deflationism and nominalism.

That truth-theoretic semantics does not fit well with nominalism should not be surprising, since it has been common to argue for Platonism precisely by considering what follows from the truth of various sentences. The most common argument is that since "that"-clauses are singular terms, one must posit objects as referents in order to explain how sentences containing them can be true. The form of this argument is age-old, and can be found in various writings of Plato and Frege (e.g., Frege (1884: § 60)), though they mainly speak of numbers and numerals. In Frege's (1892) and (1918), however, the argument for propositions are given a clear statement. The argument, then, is that "that"-clauses and expressions like "the belief (thought, fact, etc.) that snow is white" are singular terms and all singular terms figuring in true sentences have referents. Below, I will defend the claim that "that"-clauses are singular terms, so the question is how to block the commitment to propositions. The answer is simple: "true" must not be used in the semantics at all, since (NC) requires that no expressions primarily applied to "that"-clauses be used. Therefore, this argument does not create a problem. Of course, we *intuit* that "I believe that snow is white" is true, so I believe that snow is white, so there is something that I believe. But in saying this, we are only doing semantics in the material mode, i.e., displaying what facts about our use of "that"-clauses that the semantics is to explain. (This kind of response will be more thoroughly discussed below.)

The present point is that an adherent of truth-theoretic semantics cannot make this move, and that it is not obvious how he should respond to the argument. He cannot deny the truth of all sentences with "that"-clauses without being unable to account for linguistic practice. To say this does not beg the question against such eliminativists who want to deny the truth of all these sentences, for I am only making the more uncontroversial claim that such an eliminativist cannot, or at least, cannot reasonably, adopt a truth-conditional semantics. How, then, can one admit the truth of sentences containing "that"-clauses without admitting propositions, i.e., things believed, etc.?

Denying that "that"-clauses are singular terms is a possible move here, but there is a related argument that cannot be thus blocked. The truth-conditionalist nominalist must take at least some sentences of the form "*x V*s

that *p*" to be true, in order to account for our linguistic practice. Even though he can deny that "that"-clauses are singular terms, it is very implausible to deny that an existential sentence, "*x V*s something" can be inferred from a sentence of the form "*x V*s that *p*". But on a truth-theoretic semantics, valid inferences should be truth-preserving, so the existential sentence must be true. In order to avoid commitment to propositions, i.e., that there is something that someone believes, knows, says, etc., the truth-conditionalist must then deny the homophonic T-biconditionals for sentences of the form "*x V*s something". And this, rather than denying existential generalisation, has indeed been the common reaction to the dilemma. This respects (NC), which is typically accepted by nominalists, but it forces one to hold that, e.g., "Kripke believes something" is true while denying that Kripke believes something, and so on, which is of course rather counter-intuitive. The alternative semantics coupled with this claim consists in analyzing the "logical form" or "deep structure" of the problematic sentences so as to give them truth-conditions different from those given by homophonic T-sentences. Truth-conditionalist nominalists thus typically respond to the realist arguments from the truth of sentences with "that"-clauses by denying that "that"-clauses are singular terms, and by denying T-sentences for what looks like quantifications over propositions.

It may seem, however, that the nominalist need not make the unattractive move of denying T-sentences. After all, if the analyses show how sentences like "Something is believed" and "There are things believed (known, assumed, etc.)" can be true without commitment to propositions, why should it not be possible to endorse such sentences in the semantics, thus abandoning (NC)? But given how "proposition" is used in its philosophical sense, the sentence "There are propositions" just follows from such sentences. So, this nominalist has also to admit that "There are propositions" is true. Since true sentences are endorsed on this strategy, the realist's own claim will be explicitly asserted. It therefore seems that the truth-conditionalist nominalist must either deny the T-sentences after all, or find some way of distinguishing sentences like "There are things believed (known, assumed, etc.)" from "There are propositions". But the intent of the latter simply is to say the same as the former. So, if, as seems plausible, they cannot be distinguished in the relevant way, the nominalist must deny the T-sentences after all. But there is a further implausible consequence of this. Since "There are things believed (known, assumed, etc.)" and "There are propositions" cannot be distinguished in the relevant way, the latter must be held to follow from the former. But then, "There are propositions" must be held true. However, it cannot be given

a homophonic T-sentence. When giving the semantics for this sentence, then, it must be held to diverge radically from, e.g., "There are stones".

What, then, is the positive account of paraphrase nominalists? Rudimentary varieties of this idea were vented in the middle ages, e.g., in Ockham's "theory of exposition" (see, e.g., Freddoso and Schuurman (1980: II.11)). Modern paraphrase nominalism may be seen as a syntactico-semantic analysis for a certain range of expressions in natural languages, which exploits various formal methods in support of their ontological view. On these theories, "that"-clauses are typically not regarded as proper syntactic units even on a superficial level (let alone as singular terms). Rather, a sentence "*x* *V*s that *p*" is analyzed as consisting of a left-most noun-phrase left, a right-most sentence, and the expression "*V*s that", to be regarded syntactically rather as a function from a sentence to a predicate.[1] Prior also proposed that what looks like quantification over propositions, e.g., "He says everything I believe" is to be treated as having as its "logical form" a propositionally quantified sentence, "(*p*)(If I believe that *p*, then he says that *p*". These claims must be taken to concern syntactic "deep structure" or "logical form", for clearly, this natural language sentence has the same superficial form as, e.g., "He stole everything I own". Prior's propositionally quantified paraphrase does not obviously give the wrong meaning to the target sentence. But one may wonder why the similarity (or identity, if you will) of meaning should give any evidence for a syntactic claim. The very idea of deep structure or logical forms has been hotly debated, and what is known for sure is only that it will be long before there is agreement on the number, meaning and existence of syntactic and semantic levels. It is thus far from a firm ground on which to base a nominalistic project.

Nominalists who distinguish syntactic levels do so for merely semantic reasons, why it may be difficult to see why syntax is brought up at all. Let me clarify this point by an example. A common argument against various proposals of paraphrase nominalists is that, e.g., the expression "the average *x*" seems more apt for "paraphrasing away" than the nominalist's target of paraphrase. For instance, when giving, as a paraphrase of "The average American has 2.3 children", the sentence "Americans, on average, have 2.3 children", one seems to give a plausible account of what we *mean* by the latter. Then, this is contrasted to the nominalist proposal, where, e.g., "3 is a

---

[1] This view is endorsed, more or less explicitly, by Russell (1910), (1912: XII), Quine (1960: 216), Prior (1971: 16ff.), Tye (1989), and Matthews (1994), and it receives a very elaborated treatment in Moltmann (2003).

prime" is paraphrased as "The sentence '3 is prime' follows from the axioms", which seems to ascribe the wrong meaning to the sentence. These observations are often fine as far as they go. But it may be surprising that one takes this to show that the analysis of "the average American" is a better *syntactic* analysis. Even given that the paraphrase really does give the "real meaning" of the sentence containing "the average American" (whatever that is supposed to mean), some further argument is needed in order to show that this has any relevance for syntax.

Nominalists in general, including myself, are sympathetic to the idea that it is superficial grammar that lures people into realism. But one is not required to elaborate this idea by distinguishing different syntactic layers. *A fortiori*, one need not make the quixotic claim that "that"-clauses are not syntactic units. One can take "that"-clauses to be singular terms, and simply say that not all of these refer to objects, and that realism is the result of over-generalising from referring singular terms. Propositions are avoided, and without exotic linguistic analysis, by taking certain syntactically cotypical expressions to have radically different types of semantic functioning. We know that, e.g., noun-phrases, like quantifiers and names, have radically different semantic functioning, so why not singular terms? In section 5.4, we will see how to treat "that"-clauses as singular terms without commitment to propositions. It would also be better, and more in line with current linguistics, to treat all superficially similar quantifications the same way, contrary to Prior's analysis. In section 5.5, we will show to do this in obedience of (NC). The important lesson of these sections is that this can only be done with a non-alethic semantics, if we are to avoid the consequences here highlighted: the counter-intuitive truth-conditions of various sentences, the unnatural syntactic analyses and the result that "There are propositions" is judged true.

Paraphrase nominalism has been somewhat replaced by fictionalist strategies for nominalists who endorse truth-theoretic semantics (e.g., Field (1980), (1989)). On this view, ordinary intuitions and use are supposed to be accounted for by the positive value of *truth in a fiction*. Thus, the sentence "1+1=2" is not literally true, but only true in the useful fiction of mathematics, just as "Sherlock Holmes is a detective" is not literally true, but true in the novels by Conan A. Doyle. This is supposed to avoid both the syntactic claims of paraphrase nominalism and the strange consequences of non-factualist theories such as expressivism. It is not obvious just how "truth-theoretic" such semantics is, however. In particular, one may wonder how much would be lost by rephrasing away "true". From the claim that a sentence

is true in the fiction (as opposed to just true), the sentence should not itself follow, if nominalism is to be defended. What intuitively follows from "'*p*' is true in the fiction", however, is "In the fiction, *p*". Could one not then simply replace "*s* is true in the fiction" with, "*s* is affirmed by the fiction", or some such? It may be also that "true" only plays a purely disquotational role here, i.e., that it is used only to express the infinitely many instances of "In the fiction, *p*", where "*p*" is to be replaced by appropriate mathematical sentences. If so, then the semantics will not be truth-theoretic. The fictionalist will then rather explain mathematics by providing a theory of how mathematicians create a fiction and how things can be or become the case in this fiction.

It is not obvious how much similarity can plausibly be claimed between mathematics and the paradigmatic cases of fictional truth besides the negative property of not being literally true. Proving mathematical theorems is quite unlike creating or consuming narratives (writing or reading novels, etc.), which are the known ways of determining what is true in a fiction. What is true in a fiction is determined by *fiat* and with complete freedom by an individual author or artist, while mathematicians are obliged to operate within certain bounds, e.g., respecting the law of non-contradiction (Katz (1998: 12ff.)) and requiring evidence of some kind (Resnik (1997: 188f.)). I think such considerations shows that the appeal to fiction is more charitably interpreted as a gesture toward the view that the acceptance of the sentences in question are *not* to be accounted for in terms of their *truth proper*, but *some other way*. The positive theory in terms of fiction seems possible to replace for an explicitly non-alethic theory without great loss for a nominalist. In the actual positive accounts given by fictionalists, acceptable mathematical sentences seem much to be called "true in fiction" in name only, and the positive similarities with paradigmatic cases of truth in fiction are downplayed. It is of course necessary to deny many similarities for the theory to avoid gross implausibility. But when the obvious differences between paradigmatic fiction and mathematics have been dealt with, mostly negative features remain. Therefore, I do not see that there should be any loss at all, given the fictionalists' general (pre-semantic) *desiderata*, if one were to replace "*s* is true in the fiction of mathematics" with something like "the rules of mathematics legitimate (or mandate) the acceptance of *s*". They seem equally well suited to explain linguistic behaviour, and differ only in that the fictionalist account claims some connection to paradigmatic fiction. If so, then this goes to support the original point: that from a nominalist perspective, some non-alethic property is better suited to account for the discourse.

As concerns propositions specifically, the fictionalist would presumably speak of sentences with "that"-clauses being true in the fiction of folk-psychology and folk-semantics. Here, there is an element that does not arise in the case of mathematics: speakers' correct judgments vary depending on empirical circumstances. Presumably, folk-psychology is here taken to be a theory which states relations between behaviour and propositional attitudes. Here, there seems again to be no obvious reason why we should claim that the sentences licensed by this theory (given empirical evidence about behaviour) should be called "true (given the behaviour) in the fiction of folk-psychology", rather than merely "licensed (given the behaviour) by the rules of folk-psychology". Further, a fictionalist about propositions will not, of course, believe that there is a reduction of propositional attitudes to behaviour, i.e., necessary equivalences relating them. Rather, it should be possible for one ascription of propositional attitudes to be licensed by the facts about behaviour at one moment, and an incompatible sentence to be licensed later, as new facts about behaviour have emerged. Either, then, contradictions are taken as true in this fiction, or what is true is taken to change over time. This predicament is common for justification, of course, but one hesitates to say that the truth-value changes, or that the sentence in such a situation is both true and false. If so, then a normative, non-alethic positive property of sentences would be better to appeal to than truth in a fiction.

We will now look closer at the nominalist response to the realist's argument from intuition, which is that the intuition itself is simply a fact of usage, which is to be explained by a nominalist semantics. (A truth-conditionalist can also make this move, but, as we have seen, this entails paraphrase nominalism.) The final, nominalist semantics must ascribe a positive semantic value to the sentences that speakers assent to (and that nominalists assent to when doing semantics in the material mode), but which (NC) forbids us to use in the semantics. *A fortiori*, it will claim that the sentence "There are propositions (i.e., things one may believe, know, etc.)" is analytic, because it is an inference from analytic sentences like "That bachelors are unmarried is true". How, then, is "analytic" to be understood? I intend to use it, in a use-theoretic spirit, as cognate to the notion of inference that has been used throughout. Intuitively, if B can be inferred from A, then the assertibility of A guarantees that of B. That is, if A is assertible, no further sentence needs to be assertible for B to be. This much simply lies in the notion of inference. Also the truth-conditionalist grants a corresponding guarantee, but speaks instead of a sentence's truth guaranteeing that of another. On the

present account, an analytic sentence is then just a sentence which is guaranteed to be assertible independently of what other sentences are assertible. For this use of "analytic" to defeat the strategy, it must somehow be shown that the claim that "*p*" is analytic entails that *p*. For then, the claim that "There are propositions" is analytic (and that various sentences with "that"-clauses are analytic) entails that there are propositions. To show that this holds, one must show that any elucidation of analyticity or inference available to the deflationist must have this consequence. But this does not hold for the candidates actually adopted by various deflationists and "use-theorists". For instance, the claim that speakers unconditionally assent to "*p*" does not entail that *p*, nor does the claim that a sentence "*p*" is assertible, on many elucidations of this notion. For instance, that speakers reinforce the utterance of or assent to "*p*" (cf. 5.7) does not have the consequence that *p*, nor does the claim that there is a verification for "*p*". Since typical use-theoretic proposals for elucidating analyticity do not have this feature, I do not see that there is any risk that calling "There are propositions" analytic will violate (NC). In particular, given that the step from "*p*" to "That *p* is true" is merely stipulated as unconditionally assertibility-preserving, the consequences of the claim that the latter is assertible will be identical with the consequences of the claim that the former is. So, the assertibility conditions of "Snow is white" are identical to those of "That snow is white is true". If the claim that the former is assertible does not entail the existence of abstract objects, nor will the claim that the latter is assertible. Similarly, as we will see in 5.6, the derivation of "There are propositions" from any sentence containing a "that"-clause is taken as unconditionally assertibility-preserving by the stipulation of "proposition".

Accepting that sentences with "that"-clauses entail existential sentences is similarly to be explained as consisting in the fact that the assertibility of any sentence of the form "*F*(that *p*)" guarantees the assertibility of the corresponding "*F*(something)", which is a consequence of the fact that "that"-clauses are introduced as singular terms. The apparent connection between truth and validity is again taken as fact of usage, as in 3.4, where this intuition was explained by reference to the deflationary theory of "true". Similarly, our intuition that the sentence "We believe things" is *true*, is to be treated as a fact of usage, namely the fact that we are strongly disposed to assent to the sentence "'We believe things' is true", which in turn is to be understood by (D), (ST), and other linguistic facts. The same move is made in response to the realist argument from singular terms, that "Kripke believes that snow is

white" is true, so there must be some thing referred to by the term, which is *what* Kripke believes. To object that there is more to account for than the correct assent to these sentences, namely that we really seem to believe things, and that it the sentence about Kripke really is true and entails that there are propositions, is to require, unreasonably, that the nominalist is to take the existence of propositions as a datum. Our (unconditional) assent to the relevant sentences is the only uncontroversial fact, and so, the only one that needs explaining.

A successful semantics which satisfies (NC) may be regarded as required by nominalism, but not something that furthers its case positively. But I believe a successful nominalist semantics also has the effect of neutralizing realist intuitions which would tip the balance in favour of nominalism. For *if* the semantic theory succeeds in explaining the uncontroversial facts about our use of "that"-clauses and expressions that take them as objects (attitude verbs, "true", etc.), then it will also explain why the realists' very natural questions and objections make sense in the first place. Among the *explanantia* here will be such facts as the validity of existential generalisation and that we have a tendency to ask, questions of the form "What are *F*s?" when we have a meaningful general term, and so on. We can see already that relevant questions like "What are these things we believe?" and "Do they exist?" are clearly well-formed, and that they are of the same form as various questions which *do* have interesting and informative answers, for instance, "What are stars?", "Does phlogiston exist or not?". So, *if* the semantic theory succeeds, we will also have explained why these questions seem fair and legitimate in the first place, since it is simply a linguistic datum that they are. However, these questions will in the end be explained as on a par with, "But who is this average person?", only less obviously so. Thus, if a well-formed question cannot be straightforwardly answered, this fact should itself be satisfyingly explained by the theory. The most important linguistic account of realist intuitions, however, will derive from the claim that "that"-clauses are singular terms, and the intralinguistic, inferential characterization of this notion. Singular terms are essentially such as to make it *prima facie* reasonable to take them to refer. Existential generalisation, for instance, must be taken either as a criterion or a symptom of being a singular term, and this inference step is also a reason why we should wonder what *thing* is it that is said to be believed, known, etc.

That way, the execution of the nominalistic semantic project will also be a polemic move in the ontological debate. The intuition that there clearly seem

to be propositions, and that they are sometimes true, is the kind of thing realists will take to be arguments in their favour. But if these phenomena are regarded as pieces of linguistic data that can be explained in obedience of (NC), then the rug is in a way pulled from under the realist. We agree that what the realists accepts, like "There are true propositions", is analytic. But since analyticity does not entail truth, we are not defeated by the obvious plausibility of the realist's assertions. Common sense is not violated, since the utterances are counted as correct in the same sense as other unproblematic utterances, e.g., of "Snow is white", since the sense of "correct" that is used in the semantics is plausibly the same for all sentences. The difference, as noted above, will only concern the type of conditions for the utterance to be correct. With "Snow is white", the referent of "snow" (namely, snow) may be appealed to in stating the correctness conditions of "Snow is white", but nothing that can be taken as the referent of a "that"-clause will be appealed to in stating the correctness conditions for sentences containing it. Likewise, the correctness-conditions of sentences containing "white" (as opposed to "true") may be stated by using "white" (more on this in 5.7). Even if the semantic content of "white" were given entirely without such mention, but, e.g., by reference to a private *quale* (which I find very implausible), there will still be a causal story relating the word to white things (*via* the *quale*). But there will be no causal story relating "true" to true things or "that"-clauses to their alleged referents. Even on such a theory, "true" would be distinguished from "white" in that the semantic account of "white" would not be exhaustive of whiteness. The idea, then, is that if the realist's utterances can be explained consistently with the denial of propositions, then they no longer constitute an argument for realism. The trick is to explain linguistic behaviour in non-representational terms, and to take the plausibility of the realist's assertions as just another fact of usage. The final arbiter, here as everywhere, should be the dispersion of mystery, but the impossibility of giving a straight answer to a question is no mystery if it can be predicted by an otherwise satisfactory theory. The only possible refutation of this view, as far as I can see, would consist in showing that there is some uncontroversial datum that cannot be explained without commitment to propositions. But this argument cannot be anything like "Propositions exist, since it is a truism that people believe things, etc.".

Above, I intended to show that nominalism can be made more plausible by combining it with a non-alethic semantics, which, in turn, is mandated by deflationism. But it also seems that the kind of deflationism defended above

requires nominalism. Or, at least, it is not *plausibly combined* with the view that "that"-clauses refer to objects (i.e., propositions), even if it may be difficult to find a conclusive argument that it *excludes* it. Consider (D), which is meant to give an exhaustive semantic account of "true". Now suppose that "that"-clauses refer to propositions. It seems reasonable that when a predicate applies primarily to a certain type of term, and these terms refer to a certain kind of object, the semantic account of the predicate should somehow mention these objects. For present purposes, if "that"-clauses refer to propositions, then "true" should be taken to describe these objects. But since (D) is meant to give an *exhaustive* semantic characterization of "true", it is hard to see how it can explain what connection "true" has with propositions at all. Principle (D) is wholly intralinguistic, so any alleged referent of a "that"-clause would play no role in accounting for sentences of the form "That p is true". For these are just to be considered as equivalent to the corresponding "p", and that, together with facts about other expressions, is meant to explain all (semantically relevant) facts about our use of "true". In all other cases of referring terms, one takes sentences containing them to say something about the object, to have truth- or assertibility-conditions depending on what is the case concerning the object, or some other way make essential use of the referent. Even when the predicate is given a non-factualist analysis, as perhaps the predicate in "a is beautiful", the semantic account will still mention the object. One might hold, for instance, that this sentence is used to aesthetically commend the referent of the term "a", or some such. If "that"-clauses have referents, holding that (D) is exhaustive of "true" would make sentences of the form "That p is true" a somewhat mysterious exception.

The idea may be clarified by focussing more on speakers' semantic competence. If "that"-clauses refer to propositions, competent users of "that"-clauses must have some grasp of propositions. But if "true" primarily applies to "that"-clauses, then competent users of "true", it seems, must also have some grasp of what is said about a proposition when a sentence of the form "That p is true" is uttered. This seems trivially to be true in uncontroversial cases of referring terms. Competent users of "red" obviously have some grasp of what an object, referred to by a term t, must be like when "t is red" is to be accepted, or is true. They know what the sentence says about the object, and so on. But the claim that (D) gives an exhaustive semantic account of "true" leaves no room for any explanation of how "true" relates to propositions. This is reasonably something competent users of "true" should know something about, if "that"-clauses refer to propositions. Since it was shown in Chapter 3

that no deflationary theory can be properly formulated unless it gives a purely metalinguistic claim, it seems that any properly formulated deflationism will be in tension with the view that the terms to which "is true" primarily applies refer to objects. Any such theory must be wholly intralinguistic, and is supposed to give an exhaustive semantic account, so how can it possibly do justice to the idea that "is true" relates in some semantically essential way to the objects?

If "that"-clauses are not taken to refer to anything, however, the mystery disappears. "That"-clauses are regarded as term-like expressions, following the subject-predicate form that natural languages love so much, which must be coupled with predicates in order that proper sentences be formed. Given a semantic account like (D), further, it can be shown to enable the expressive strengthening witnessed in 1.4, and the rest of the story was given in Chapter 4. This story may be doubted, of course, but the present point is only that nominalism with respect to "that"-clauses avoids a mystery that deflationists would have to explain if "that"-clauses are taken to be referring expressions.

5.4 THE INTRALINGUISTIC NOTION OF SINGULAR TERM

In this section, I will argue that "that"-clauses are to be seen as syntactic units, and, more precisely, as singular terms. We may cautiously note, though, that most of the semantic accounts of various occurrences of "true" in 4.3 do not need the characterization of "that"-clauses as singular terms, but only the uncontroversial observation that "that"-clauses behave in some important ways *as* paradigmatic singular terms. Those resemblances were exploited in that section, but one need not agree that there are any further common denominators in order to accept those explanations. But fair is fair. We will here also see that the notion of singular term is a bipartite notion with a syntactic component and a logico-semantic one. The discussion of the claim that "that"-clauses are singular terms will accordingly be split in two, and we will begin with the syntactic part.

The syntactic "component" of the notion of singular term is the condition that any singular term is a syntactic unit, and, more precisely, a noun-phrase. We may give an intuitive explanation of the notion of a unit by giving examples of what is reasonably *not* a syntactic unit. In the sentence preceding

this one, the sequence "notion by giving" is not reasonably taken as a unit. How exactly to define the notion may be debatable, but it seems fairly clear that it has to do with which categories to use when stating grammatical rules or generalisations. It may be that in a final, "ideal" theory of grammar, ordinary grammatical categories like "noun phrase" will not be used, just as a final physics will probably not use the category of cups, teeth, wood, etc. But traditional syntactic notions will still be seen, from the perspective of the final theory, to own some reality, which will justify the use of these "higher-level", vaguer but pragmatically indispensable, categories. That is, the idea that stating generalisations using "noun phrase", rather than some term covering "notion by giving", as used in the above sentence, will surely be justified by the ideal final theory of grammar in that it yields *better* generalisations, even if they are not perfect like those in the ideal, final theory.

Now, there is plenty of intuitive and "naïve" evidence that "that"-clauses are syntactic units, and traditional "school grammars" accordingly categorize them and other "sentential" expression that behave as noun phrases as *nominal clauses*. For instance, the simplest and most typical kind of occurrence of a "that"-clause is in a sentence which seems superficially to combine it with a one-place predicate. This may be a copula and an adjective ("is true"), a noun phrase and a transitive verb ("He believes"), and so on. School grammars usually do not say that they *are* noun phrases, however, since these are defined as having a noun as its head. However, this seems like an unmotivated complication. Instead, cotypicality should be determined by external relations to other expressions, and having a noun as head should be considered just as irrelevant for syntax as, e.g., containing a certain number of words.

Taking "that"-clauses as noun-phrases will license sentences like "He believes the Eiffel Tower" as grammatical, of course, but traditional grammars also license "He drank procrastination", etc. These sentences are semantically anomalous, not ungrammatical. The idea that they are ungrammatical seems to stem from two further ideas. First, the common but dubious idea that synonymous sentences must somehow share syntactic structure (and *vice versa*). Secondly, the idea that these semantically anomalous sentences are "meaningless". It is not obvious that they are, however. They might simply be taken as (trivially) false, hence meaningful. The main problem with this idea, however, is that "meaning" and "meaningfulness" are too coarse-grained to be applied thus uncritically. There is an obvious difference between "He believes the Eiffel Tower" and "Gag rath

blayt" or "Run with sits there". The narrower notion of semantic anomaly better characterizes "He believes the Eiffel Tower", and the contrast between this sentence and the other incorrect sentences should make it clearer that it should not be called ungrammatical. What may seem worse, however, is that this account also has to allow as grammatical

(?)　　　John is to the left of that snow is white.[2]

Intuitions are unreliable here, however. As explained in note 4, Chapter 4, there are "centre-branching" sentences which appear ungrammatical, like "The man the boy the girl kissed met ran", which are grammatical, but are constructed according to the rules in such a way that we are unable to parse them correctly. Given that (?) is so obviously semantically anomalous, it is not clear how to classify its incorrectness. A positive argument for taking (?) as grammatical, however, might be the intelligibility of the following. *There are things believed, e.g., that snow is white, and so on. These things may, strange as it seems, be to the left or right of people. For instance, the thing I most firmly believe, that snow is white, is to the right of John. So, that snow is white is to the right of John. That is, John is to the left of that snow is white.* Try pronouncing the last sentence with appropriate pauses and stresses. Well, it still feels a bit wrong, but the penultimate sentence does not, and this seems to force us to take (?) as grammatical after all, and explain its awkwardness in terms of the specific order of phrases.

　　In the end, there is always the retreat position of taking "that"-clauses to be merely nominal clauses. Until proven wrong, however, we take "that"-clauses to be noun phrases, on the present use of the term "noun phrase". On the traditional definition of "noun phrase", my claim is that noun phrases and "that"-clauses belong to the same category, and that a further division between them will be syntactically superfluous. The point of showing that "that"-clauses are noun-phrases is that it makes it easier to defend the claim that realists err in linguistic over-generalisation. Together with the fact that "that"-clauses are also logico-semantically singular terms, it is only understandable that one should take them to be like singular terms that are also typical traditional noun phrases, i.e., referring expressions. The more reasonable one can make the view that "that"-clauses refer, consistently with the denial with this claim, the better the chances of neutralizing realist intuitions.

---

[2] I owe this example to Peter Pagin.

What is the positive evidence, then, for the claim that "that"-clauses, *pace* Prior, are syntactic units? There is too much of it to rehearse here, but we may here provide a sample for the uninitiated. One piece of evidence comes from splitting. We do not say, *"It is true that, I presume, *p*", but, "It is true, I presume, that *p*". This, as most arguments on these issues, is a good indication but not conclusive evidence. Since many believe that "that"-clauses are singular terms, indeed, referring expressions, it is to be expected that there is more to their behaviour that indicates that they are syntactic units. There is thus the case of existential generalisation, noted above, the fact that we say "That snow is white is the most ordinary thing to say", and many further more or less hackneyed examples that need not be listed here.[3]

The syntactic condition upon singular terms should now be supplemented with a logico-semantic one, which it is more difficult to state.[4] One can provide various specific criteria, pertaining to inferential relations between sentences containing the expressions and others, but it is not clear which should be regarded as symptoms of the correct constraint, and which are to be regarded as parts of the constraint itself. One thing that intuitively indicates that "that"-clauses are singular terms is that we can derive from "Kripke believes that snow is white", the sentence "The *thing* Kripke believes is that snow is white". But this should plausibly be taken as a consequence of the definition of "singular term", together with the semantics of "thing", rather than being in the condition itself. What is important is that the constraints be intralinguistic, more precisely, inferential, rather than one that requires an extralinguistic property.

Those insisting on an extralinguistic criterion will typically take this to be reference to an object. One cannot deny that some singular terms refer to objects. This reveals a potential terminological confusion. For even a nominalist that agrees with what I have said may want to call singular terms that refer genuine singular terms, and hold "that"-clauses to be "merely

---

[3] For more interesting examples of evidence, see, e.g., Bealer (1982), (1998), Bealer and Mönnich (1989), Schiffer (1992), (1996) and Parsons (1993).

[4] There is no tension here with the above expressed view that syntax and semantics come apart. The claim was that syntactically cotypical expressions need not be semantically cotypical. I define "singular term" as something that meets both syntactic and semantic conditions, so it is not a syntactic category. This should be obvious, but confusion may be caused by the fact that I above reasoned on the assumption that if something is a singular term, it is a syntactic unit. This is of course consistent with (and in fact follows from) the claim that if something is a singular term, it is a noun-phrase.

superficial" singular terms. He may then join the project of giving an inferential account of the notion of (superficial) singular term. But I think this terminology is flawed. One reason not to make a distinction between genuine (i.e., referring) and superficial singular terms is that the question which category an expression belongs should not depend on *a posteriori*, unknown, and linguistically irrelevant facts. A linguistic theory should not require knowledge of, say, physics or history in order to be able to classify expressions (knowledge that physicists or historians may not have themselves). One may want to avoid this consequence by defining genuine singular terms as those which *purport* to refer to objects. But this will create the problem that realists will say that various expressions purport to refer (and succeed) where the nominalist denies this. Just as linguistic classification should not depend on knowledge of physics, it should not depend on philosophical opinion. Defining "singular term" in terms of reference simply does not seem appropriate.

Since we have an intuitive notion of singular term, and since it seems relevant to metaphysical issues, it is still an important question how it should be defined. But it should then be made uncontroversially applicable. That is, we should find criteria that satisfy our intuition that there is something in common among proper names, numerals, "that"-clauses, and so on, the application of which does not turn on any opinion at stake in metaphysical disputes. A good candidate is precisely a definition in terms of inferential properties. The relevant metaphysical questions, properly formulated, are thus, "What makes an expression a singular term?", "Do these necessarily refer to objects?", and "Which of them do and which do not refer?".

Incidentally, this has some relevance for the polemical situation concerning realism. Crispin Wright has argued cogently (1983: v and vii), and that there is no non-arbitrary way of distinguishing "real" (i.e., referring) singular terms from merely "stylistic" (i.e., uncontroversially non-referring) ones, save by appeal to foresworn referential notions. But most realists will probably follow me in not taking all singular terms, inferentially defined, to refer, since they would otherwise have to admit more objects than desired, such as the average American, and so on. But if there is only one legitimate notion of singular term, and the expressions falling under this notion do not all refer, then the argument that there are true sentences with singular terms does not establish that there are propositions (or other *abstracta*). But if one defines "singular term" in such a way that it follows that any singular term refers, then since there uncontroversially are other notions of singular term, e.g., the

notion of a superficial singular term, the argument will be question-begging in assuming that "that"-clauses belong to the referring kind.

It is a moot question exactly what inferential constraints should be set on singular terms. Clearly, more than the licensing of existential generalisation is needed since "something" would otherwise count as a singular term. I will not here contribute to the literature on defining "singular term" intralinguistically. There are several attempts at intralinguistic accounts of singular terms which may serve as starting points for the present nominalistic project.[5] It should be noted, however, that while this at base Fregean enterprise has mainly been intended to reveal the nature of various types of entities *via* logico-syntactic analysis, this is of course contrary to my intentions. To me, it seems too idealistic to say that the notion of object should be characterized by reference to languages (if they can be characterized at all, rather than being primitively understood). By the same token, if we are to countenance abstract objects after all, we should follow intuition and linguistic practice and take them to be eternal and wholly objective, and to explain our knowledge of and reference to them on the basis of what they are like, not the other way around. Now that we do away with them, our intuition of eternality and objectivity are instead taken as a further linguistic datum to be accommodated by the nominalist theory. This objectivity could perhaps be inferentially explained. For instance, there seem not to be any non-trivial correct inferences between sentences describing our language use and sentences containing abstract singular terms, like "1+1=2" or "That 1=1 entails that 1=something". The Frege-Dummettian middle way between Platonism and nominalism seems to me to unite the worst from both views in admitting abstract object but make them dependent on us.

Although I believe the definition of singular term should be entirely intralinguistic, a theory about singular terms need not completely lack mention of reference to objects. In biology, a phenomenon often requires a diachronic explanation to be exhaustively explained, e.g., the appendix in humans. It may be, likewise, that the category of singular term emerged, in the wake of human language, essentially as a category of expressions referring to material objects, but that non-referring expressions were subsequently introduced as singular terms, i.e., as having the same inferential properties. Perhaps something similar is true of each speaker's individual learning pro-

---

[5] See, e.g., Frege (1884: §§ 26, 60-62), Dummett (1973: 57ff.), Geach (1975), Wright (1983: ix), Hale (1987: Ch. 2), Hale and Wright (2001: Chs. 1 and 2), Brandom (1994: Ch. 6), Heck (unpublished).

cess (perhaps, to some extent, idiolectogeny (if I may) recapitulates glossogeny). This, in any case, could do justice to the intuitive connection between singular terms and reference.

5.5 OUTLINES OF A NOMINALIST SEMANTICS

We will now look at a few more specific questions about how a nominalist treatment on the above lines should deal with various sentences that superficially refer to or quantify over propositions. Simple sentences with a "that"-clause, to begin with, are given a standard (i.e., relational) syntactic analysis, but what can be said about their semantics? I believe that the semantic character of a sentence like "Kripke believes that snow is white" is a function from the semantic characters of "Kripke", "believes", and the embedded sentence "Snow is white". Some seem to take this view to motivate the syntactic analysis on which "that"-clauses are not syntactic units. But besides resting on the above-criticized idea that syntax should mirror semantics, the semantic claim in this case does not even seem to motivate the syntactic claim, even granted such a connection. For why should the "that" go with "believes", rather than with the sentence? It therefore seems more plausible to accept the semantic claim and regard "that"-clauses as syntactic units, and, hence, regard the "that" as semantically superfluous, serving only to create an expression that can act as a grammatical object to the verb.[6] It should be noted that while a realist may agree that the meaning of a sentence "x Vs that p" is a function partly from the meaning of "p", he takes this to be a consequence of the fact that the meaning of this sentence determines which proposition is referred to by the "that"-clause. The present claim is that "that" is semantically superfluous, thus contradicting the realist's analysis.

There seem to be special advantages to a view that combines the idea that (1) the "that"-clause in a belief ascription is a singular term and the view that

---

[6] Davidson has argued, surprisingly, that "we cannot *understand* [sentences of the form "the proposition that p is true iff p"] unless we can see how to make use of the same semantic features of the repeated sentence in both of its appearances–make use of them in giving the semantics of the schema instances. I do not see how this can be done." (1996: 274, original emphasis). But, surely, we *understand* them! Furthermore, the fact that we cannot (now) give a compositional semantic account of "the proposition that p" is surely no reason to say that we do not understand it.

(2) the meaning of the ascription is only a function from the subject, the verb and the sentence. To put the matter briefly, it seems that a theory which manages to combine (1) and (2) is more likely to be able to take the good and eschew the bad from both Russell's "multiple relations analysis"[7] and the "face-value", or "relational" analysis of belief ascriptions (and other ascriptions of propositional attitudes). The latter is simply the view that belief is a relation between a believer and a believed proposition, whereas on Russell's account, belief is related to differing numbers of objects that are referred to by expressions within the "that"-clause. Thus, Russell says, "when Othello judges [e.g., believes] that Desdemona loves Cassio, Othello is the subject, while the objects are Desdemona and loving and Cassio" (1912: 126). This is as yet only a very rough idea, of course, but it has attractive features.[8] This is also the kind of analysis endorsed by paraphrase nominalists of Prior's brand. My idea, then, is that the various intuitions and arguments supporting these rivalling ideas could be better accommodated by a theory combining (1) and (2), than by either of the rivals. The present account does not agree with the relational theory that "believes" should be explained as expressing a relation to propositions, of course. But this idea is only another realist intuition, which is to be accounted for by the nominalist semantics. The intuition is there taken to register the termhood of "that"-clauses. The multiple relations analysis is also considered as right in a way, namely, insofar as it gives a semantic account which does not mentions propositions. More importantly, it is more congenial to the idea that the meaning of a sentence "x Vs that p" is a function from those of "x", "Vs", and "p", taking "that" as a syntactically required, but semantically redundant expression.[9] Describing the meaning-functions for

---

[7] Besides Russell (1910), (1912: XII) and (1918: IV), similar ideas have been vented by Prior (1971: 16ff.), Tye (1989), Matthews (1994), and Moltmann (2003), i.e., the very same that accept the syntactic analysis of "x Vs that p" into "x", "Vs that" and "p".

[8] Recently, the view has been promoted by Moltmann (2003) as avoiding various difficulties that afflict the view that "that"-clauses refer to propositions. Note, then, that it is not the claim that "that"-clauses are singular terms that is unfavourably compared to the Russellian view, only the view that they refer. We will have reason to look closer at Moltmann's arguments in the next section.

[9] To say that the meaning of a sentence "x Vs that p" is a function partly from that of "p" should not necessarily be taken to mean that sentences with the same semantic value can substitute one another in the matrix "x Vs that..." without change of semantic value. If semantic value is taken to be the proposition expressed, this could perhaps be said, but this is of course ruled out by (NC). An assertibility-conditional semantics could perhaps take the assertibility *conditions* of "p" to contribute systematically to those of "x Vs that p". Or,

various propositional attitude verbs is of course a major project, that cannot be as much as begin here (in 5.7, however, we will briefly discuss how (NC) relates to this task). The same, of course, goes for the project of seeing how the above account relates to the actual arguments that have emerged in the debate between Russellians and relational theorists. I hope, in any case, that this explains, if on a very general level, how the semantic contribution of a "that"-clause is treated on this nominalist semantics, and how this enables a compromise between two important rival accounts of "that"-clauses. This, in turn, shows how the guiding idea behind paraphrase nominalism can be expressed without unnecessarily entering into discussion of syntax, deep structure, and the neural mechanisms underlying linguistic competence.

In order to respect (NC), the nominalist semantics must also give some alternative account of what looks like quantification "over propositions", i.e., quantified sentences whose instances have "that"-clauses as terms. An objectual account, of course, is both truth-theoretic and entails commitment to propositions. The obvious alternative for the nominalist is some kind of substitutional, or, in any case, *intralinguistic*, interpretation (cf. Schiffer (1987: 234ff.)). A substitutional interpretation of quantification is traditionally considered a part of the non-representational paradigm, and is also, as we shall see, more promising on this type of theory of meaning than on a truth-conditional one, thus again supporting the claim that nominalists should not be truth-conditionalists.

As a preparation for our intralinguistic account of the universal quantifier, let us neutrally call the semantic content of a term whatever should be taken to be relevant for the semantic value of sentences containing it. It may be a mere referent, descriptive contents, recognition-conditions, images before the mind's eye, or whatever the reader's latest bet may be. Secondly, I do not think universal sentences are primarily of the form "Everything is *F*", but, rather, "Every *F* is *G*", the former being a special case of the latter. Now, I propose:

---

the semantics may be yet more fine-grained, taking into account the semantic properties of the subsentential expressions in the sentence following the propositional attitude verb. These sentences of course pose major problems also for truth-theoretic semantics, and I will leave the issue here.

(UQ)   A sentence "Every *F* is *G*" is assertible iff, for any term *t*, it holds, irrespectively of *t*'s semantic content, that given that "*F(t)*" is assertible, so is "*G(t)*".[10]

The idea here is that to assess the assertibility of a universal sentence, the only feature of the instantiating expressions that may be appealed to is that they are terms, i.e., their logico-syntactic identity, not what semantically distinguishes them from other terms. Secondly, the assertibility of "*F(t)*" must suffice, irrespectively of *t*'s semantic content, for the assertibility of "*G(t)*", regardless of whether "*F(t)*" or "*G(t)*" are in fact assertible. Thirdly, (UQ) is to hold also when a universal sentence is assertible due to experience. This entails that assertibility is relative to time and speaker, but this I take to be in the concept in any case. One might say that a universal sentence is assertible only due to the contents of the predicates involved and (sometimes) experience. This means that if, for *some* term *t*, the assertibility of "*F(t)*" is sufficient for that of "*G(t)*", irrespective of the semantic content of *t*, then this holds for *every* term. Note also that the adverbial phrase "irrespectively of the content of the term" in the right-hand side of (UQ) has the whole conditional in its scope. Therefore, the right-hand side can be true although "*G(t)*" is assertible partly due to the semantic content of *t*.

Let us now take the example of "Everyone in this room is blond". Suppose *a* and *b* are the only persons in "the room" and are both blond, and that *a*'s experience is indeed sufficient to make the universal sentence assertible. The idea is that, regardless of the semantic content of "*c*", *a*'s experience and the assertibility of "*c* is in the room" will be enough to make "*c* is blond" assertible, irrespective of the semantic content of "*c*". (It will also be sufficient to make "*c* is either *a* or *b*" assertible, but this is not a necessary intermediate step.) Otherwise put, given the experience and given that "*c* is in the room" is assertible, there is here nothing about the term "*c*" *other than its being a term*, due to which "*c* is blond". We have an unproblematic case when neither "*c* is *a*" or "*c* is *b*", nor their negations are assertible. This may be so when "*c*" is, say, "the person in the room who blinked exactly one year ago".

Trouble may seem to arise, however, when we consider the instance "*b* is in the room", where the term "*b*" refers to *b* and *a* knows it. In this case, both "*b* is in the room" and "*b* is blond" will be assertible partly due to the semantic content of "*b*". Again, this is in no contradiction with the right-hand

---

[10] For simplicity, I am being sloppy about the variable "*t*" in quote-contexts.

side of (UQ), since the "irrespectively" phrase takes the whole conditional in its scope. But, one might object, the grounds for the assertibility of "*b* is in the room" are not in themselves grounds for the assertibility of "*b* is blond". But this only means that the contents of the predicates do not themselves suffice for the universal sentence to be assertible. That is, experience is needed to go from "*b* is in the room" to "*b* is blond" if this is to be done without assumptions of the content of "*b*". But, the objection might continue, knowledge that "*b* is in the room" is assertible does not suffice for the assertibility of "*b* is blond" irrespectively of the content of "*b*", since *a* needs to know something about the referent of "*b*". Indeed, *a* needs certain perceptual experiences of everyone in the room, including the referent of "*b*". But *a* need not know what "*b*" refers to. There is a *de re-de dicto* ambiguity here. Speaker *a* must have certain experiences of the person who happens to be the referent of "*b*" (and of other persons in the room), but in order to conclude that "*b* is blond" is assertible from the assumption that "*b* is in the room" is assertible (plus the experiential evidence), the information that "*b*" refers to a certain individual is not necessary.

It might also be thought that (UQ) will give the wrong result in a case where a speaker has only the terms "*a*" and "*b*" in her vocabulary and there is a third person in the room, *d*, who has dark hair. But the assertibility of "*a* is blond" and "*b* is blond" will not suffice to make "Everyone in the room is blond" assertible in this case, for neither "*a* is blond" nor "*b* is blond" will be assertible irrespective of the semantic contents of "*a*" and "*b*". In this alleged counter-example, "*b* is blond" is assertible partly due to the semantic association between "*b*" and *b* (or, if you will, between "*b*" and its recognition-conditions, etc.). It could not have been assertible irrespective of semantic content, since, in the case as imagined, the perception of *d* would bar one from going from the assertibility of "*b* is in the room" to "*b* is blond", if this step is to be assertibility preserving irrespective of the content of "*b*".

There may also seem to be trouble when "*c*" is such that "*c* is not *a*" is strongly assertible. For instance, what are we to say about the above case when "*c*" is "the largest galaxy"? It may seem awkward to discuss what follows from the assertibility of "the largest galaxy is a person in the room", since this is so obviously non-assertible. But speakers are not supposed to make explicit assumptions in the way this objection assumes. The assertibility of universal sentences simply coincides with the conditions on the right-hand side of (UQ) – it in no way entails that speakers explicitly assume a given sentence. Principle (UQ) only states that the determination of the assertibility

of a given universal sentence goes by taking experience and the contents of various expressions as input, and produces a universal sentence if no information about the contents of terms is needed to go from "$F(t)$" to "$G(t)$". We may metaphorically describe the process going on as one where a semantic *homunculus* assumes the sentence "the largest galaxy is a person in the room", makes himself forget the meaning of "the largest galaxy", and then consults the experiential evidence to determine the assertibility of "the largest galaxy is blond". The fact that only one such case needs to be determined in order to determine the assertibility of a universal sentence also adds some psychological plausibility to (UQ). In the above case, where we needed to distinguish between, on the one hand, "a is blond" being assertible partly due to the semantic content of "a", and, on the other, its being assertible due to the assertibility of "a is in the room" irrespective of the semantic content, should not be taken as too subtle to be psychologically plausible. For this distinction is supposed to be made tacitly, or computationally, and so does not require that speakers actually manage to disregard the meaning of a term.

Induction may be thought to present a problem for (UQ). An induction, it may be thought, is a step from the assertibility of instances "$F(a)$ & $G(a)$", "$F(b)$ & $G(b)$", ..., and further conditions C (to be spelled out by the philosopher of science), to the assertibility of a universal sentence "Every F is G". But here, do not the particular instances matter for the assertibility of the universal sentence? If they do, then we can instead characterise induction as a step from the assertibility of instances "$F(a)$ & $G(a)$", "$F(b)$ & $G(b)$", ..., and C, to the assertibility of any sentence of the form "$G(t)$" given that of "$F(t)$", irrespective of the content of t. It need not even be thus sententially characterized at all, but can be taken to be a case where a certain type of series of experiences licenses one to infer from any sentence "$F(t)$" to a sentence "$G(t)$", irrespective of the content of t (that is, given only the contents of "F" and "G"). On both these characterizations of induction, of course, (UQ) entails that "Every F is G" is assertible in the case described.

Mark Lance (1996) gives a substitutional interpretation in much the same spirit, but he formulates the conditions in terms of the assertibility of sentences with arbitrary names. These are like the lower-case italic "a" in the assumption-sentence "Let a be a natural number...". This is similar to (UQ), since arbitrary names precisely lack semantic content in our sense, and only has a grammatical identity. But it seems that we should try to avoid the consequence that any speaker who uses a universal quantifier must operate with arbitrary names. It seems quite contentious to say that earlier humans

have used arbitrary names as long as they have used universal quantifiers. Even if this operation is taken to be tacit, (UQ) seems still less contentious since it only requires that the semantic content of universal sentences is sensitive to the semantic contents of its instances (namely, by disregarding that of the terms therein).

Of course, (UQ) also has affinities with the introduction and elimination rules of the universal quantifier in natural deduction. It is difficult to see how this rule should be applied to cover cases where a universal sentence is assertible due to empirical evidence, however. This rule might be "safer" to follow explicitly in practice, since it guarantees that we are not tacitly sneaking in assumptions depending on the content of the term used for deriving a universal sentence. But this seems to have little relevance for the semantics of quantifiers. There is another important advantage that (UQ) has over an account closer to the introduction rule of the universal quantifier in natural deduction. Take, for instance, a rule that allows one to infer a universal sentence if an instance is *analytic* and depends on no hypotheses containing the term. But for any analytic sentence "p", the corresponding "That p is true" will likewise be analytic, by (D). But then, such an introduction rule would allow us to infer "Everything is true". With the constraint of content-independence, however, this move is barred by the fact that this instance is assertible because the sentence "p" is analytic, part of the content of the term, i.e., the "that"-clause.

Here is now a further reason why nominalism, since it requires substitutional quantification, is not happily wedded to truth-conditional semantics: (UQ) would not be intelligible if "assertible" were replaced with "true", for it does not make sense to say that a sentence of the form "$F(t)$" is true, given the truth of such and such a sentence (and excluding experience, of course), irrespective of the semantic content of t. A sentence simply cannot be true irrespective of the semantic content of its parts. A speaker may however be *justified in holding* a sentence true irrespective of this, and this is surely related to the corresponding fact about assertibility. It seems that the truth-conditionalist needs to give truth-conditions of universal sentences by quantifying over every possible instance, including all non-actual instances.

There is also a cogent argument by Tomberlin (1997) against the truth-conditional substitutional interpretation of the universal quantifier. What, he asks, is to prevent us from inferring from the claim that "Everything is F" is true, that "F(the unique non-F)" is true? He takes the only possible solution to be to restrict the substituends to *referring* terms, which would conflict with

the aims of the substitutional interpretation. *A fortiori*, this would not work for a nominalist, who must take all the relevant terms to be non-referring. For our purposes, the problem as formulated above does not arise. But something analogous might. Take "the unique non-self-identical". It seems that the sentence "Everything is self-identical" cannot be assertible, since for an instance to assertible, it must be assured that the term is not like "the unique non-self-identical", since no sentence with such a term can be assertible. But this is to make an assumption about the term's content. I said earlier that sentences of the form "Everything is *F*" should be treated as special cases of sentences of the form "Every *F* is *G*". But they are special cases so common that the quantifier expression has merged in writing with the noun "thing". For a sentence of this form to be assertible, on (UQ), it must be the case that given the assertibility of "*t* is a thing", "*t* is self-identical" must be assertible irrespective of the content of *t*. But now we see that "the unique non-self-identical" creates no problem since the sentence "The unique non-self-iden-tical is a thing" can never be assertible. True, in other languages, there are universal sentences that translate "Everything is self-identical" without a corresponding noun, e.g., the German "Alles ist selbstidentisch". But since this is not of the same syntactic form as the German equivalent of "Every man is mortal", it cannot be used as a counter-example to (the German equivalent of) (UQ). It must be given a semantic analysis of its own, and therefore, it is not *ad hoc* to give it an account which avoids problems with "the unique non-self-identical". That account would simply say that a sentence "Alles ist *F*" is assertible iff for any term *t*, if (the German equivalent of) "Something is *t*" is assertible, then, irrespective of the content of *t*, "*t* ist *F*" is assertible.

Just as the above treatment of sentences of the form "*x V*s that *p*" provides a middle way between Russell's multiple relations theory (endorsed by paraphrase nominalists) and the relational analysis (endorsed by realists), the present account of the universal quantifier, enables a middle way between the syntactic analysis of Prior (in terms of propositional quantification) and an objectual interpretation (endorsed by realists). It agrees with the former that a quantified sentence like "He says everything I believe" does little more, semantically, than "$\forall p$(If I believe that *p*, then he says that *p*)", namely to license inferences to and from the instances of "If I believe that *p*, then he says that *p*". Yet it agrees with the realist's account that the sentence is both syntactically and semantically cotypical with other universally quantified sentences.

We can now also show that the problem Gupta found with Horwich's theory (cf. 1.5 and 2.12) can be solved in a simple way. The problem was that of explaining universal facts such as the fact that everything known is true. As detailed in 3.4, this is taken on the linguistic account of "true" to be a matter of explaining the (unconditional) assertibility, or analyticity, of "Everything known is true". Terms, like "England", that form semantically anomalous instances fall out as irrelevant since such instances can never be assertible. We suppose, therefore, that an arbitrary sentence of the form "That p is known" is assertible. Then, thanks to the factivity of "know", we see that if this sentence is assertible, then so is "p", irrespective of the content of the "that"-clause. This is just an instance of the inference from "x knows that p" to the corresponding "p", which is part of, or a consequence of, the semantics of "know". This rule, further, does not need to be formulated in terms of "true" and can be followed by a speaker whose language does not contain "true" or any synonym. Now, given (D), the corresponding "That p is true" must unconditionally be assertible, too. This is independent of the "that"-clause we use, since (D) does not concern any "that"-clause in particular. In this case, then, the sentence "That p is true" is assertible due to no other fact about the "that"-clause than that "That p is known" is assertible. As noted above, if for one term t, the assertibility of "t is known" suffices for that of "t is true" irrespective of the content of t, then this holds for any term. By (UQ), then, "Everything known is true" is (unconditionally) assertible.

A corresponding proof is not available for Horwich, since he would need to appeal to a specific fact about a specific proposition, one of the propositions in the Minimal Theory, whence the derivation would not be independent of the specific proposition chosen. His own proposal (cf. 2.12) may be considered less obvious, in that the rule requires that the person inferring must know infinitely many propositions, and also have second-order knowledge about this knowledge. What is worse, it seems that Horwich cannot explain the above fact without assuming that the theory of knowledge is infinite in the way (MT) is. For the explanation of the fact that everything known is true may not go by assuming that the theory of knowledge contains the claim that everything known is true. This is because he has promised to be able to explain every fact about truth with only (MT) and facts that do not pertain to truth. The fact that everything known is true is of course a fact "pertaining to truth", in the relevant sense. How, then, can Horwich explain this fact without assuming that the theory of knowledge contains (*inter alia*) every proposition of the form "If one knows that p, then p"? One idea would

be to let the theory of knowledge contain the propositionally quantified claim that (p)(If someone knows that p, then p). But given the possible ways of elucidating this claim, as detailed in 3.3, this seems merely to disguise the problem. It may not be explained in terms of truth, and it seems odd to take it to be inferentially explained, since one could then equally say that the theory of knowledge contains a rule licensing inferences from "x knows that p" to "p". There is no good way of expressing this in ordinary English, and the infinitistic interpretation is of course what is to be avoided. In any case, if this claim can be in the theory of knowledge, it should also be possible to use a propositional quantifier in the theory of truth. The upshot is that the infinity of (MT) will contaminate other theories, since they may not be formulated in terms of truth. On the present proposal, by contrast, we can give the finite claim that the inference from "x knows that p" to "p" characterizes "know", which is clearly a claim independent from (D).

Horwich's theory also faces a certain dilemma which we can avoid: in order for his theory to be general and not only concern propositions we can actually express, he must either posit a structure (or "form") of propositions or speak of possible, non-actual sentences. In (D), by contrast, we need not quantify over non-actual sentences. Principle (D) can of course remain true of English or of a given speaker's idiolect no matter which new expressions are introduced. If new expressions are introduced in the language, the new sentences formed with them will be subject to (D) just as the old ones. So, (D) is in force in a language (or idiolect) at t if it applies to every sentence that can be formed at t. When a speaker learns "true", she begins to speak in accordance with (D), and this will allow her to apply "true" to any new "that"-clause she may be able to formulate. That is, once in force, (D) will continue to be in force provided nothing special happens (like a reinterpretation of "true", amnesia, etc.). Letting (D) quantify over possible sentences would only have the superfluous intent of allowing speakers to apply it to non-actual sentences. Horwich's problem is that he needs to speak of every proposition, whether expressible or not. We only speak of a form of linguistic expression, "that"-clauses.

5.6 THE PHILOSOPHICAL NOTION OF "PROPOSITION"

It is now time to look closer at the relation between "that"-clauses and the word "proposition". Whereas in natural English, the latter usually means something like "proposal", I have been using the latter in the more encompassing, philosophical way. The reason the more general sortal term does not exist in natural languages seems to be that no restriction on the quantifier in, e.g., "Everything he believes is true" is needed. Since terms other than "that"-clauses and appropriate descriptions ("what he believes") create semantically anomalous instances (like "He believes the Eiffel Tower"), it would be superfluous to have a restriction on the quantifier, as in "Every *proposition* he believes ...".

How should the philosophical use of "proposition" be characterized consistently with our nominalism? Note that this is not an empirical matter of how to explain speakers' use of the expression, but a stipulative one. I would like to suggest that we introduce the expression "proposition" by declaring the instances of the schema

(P)   That *p* is a proposition

to be analytic. "Proposition" is thus defined as hypernymous to "that"-clauses, just as "number" is hypernymous to numerals. It has been alleged that applying "proposition" thus generally creates various semantic and syntactic anomalies. For instance, though "*x* fears a proposition" is intelligible (if odd), it should not, as (P) entails, be a consequence of "*x* fears that it will rain". Furthermore, the sentence "*x* says a proposition" just seems awkward. Strangely, however, these problems do not seem to arise if "knows" replaces "fears" and "says".[11] I have already contended that my use of "proposition" is technical and need not, therefore, do justice to intuitions the way analyses of ordinary expressions must. However, this does not seem to be an ordinary kind of stipulation, since stipulations normally do not provoke objections from awkwardness in this way.

I would like to suggest something of a trivialization of this problem. Moltmann notes that speakers' intuitions differ rather widely as to the

---

[11] These issues have been discussed by cf. Prior (1971: Ch. 2), Asher (1987), Bach (1997), King (2002), Moltmann (2003) and Schiffer (forthcoming).

meaning and intelligibility (and even the grammaticality) of various relevant sentences with "proposition", and sometimes differ depending on their degree of philosophical training (2003: 90f.). Some even refuse to take "*x* knows a proposition" to be a consequence of "*x* knows that *p*", taking it to be intelligible only on the reading, "*x* is acquainted with a proposition". I believe the trivial reason most philosophers accept this inference, but not the ones with "fears", "says", etc., is that they are *familiar* with it. We do naturally speak of "things one may know", and, epistemology being such a common topic, the broadened use of "proposition" has gradually grown upon us. Had fear been as commonly discussed as knowledge, the same would have been the case with this word. Those philosophers who know some French (but do not philosophize in it the language too often) will note that "Je sais une proposition" sounds more odd than "I know a proposition". The reason some propositional attitude verbs seem especially recalcitrant to (P), like "say", is probably because they differ from "know", "fear", etc., in not allowing terms in general as grammatical objects. One can know or fear anything, in the sense of being afraid of or acquainted with, but one cannot say the Eiffel Tower, Winston Churchill, and so on. With a little open-mindedness, however, I think anyone can "feel" the sense of "*x* fears a proposition" as an entailment of "*x* fears that it will rain". After all, we can say that when one fears or says, that *p*, that *q*, etc., it follows that there are *things* one fears and says, namely, that *p*, that *q*. But then, there should be no additional problem with calling these things "propositions". Therefore, in summary, I think we should use (P) as a definition of our technical us of "proposition" in philosophy and, simply, get used to it!

There remain certain special problematic cases, however, that we will discuss before closing this investigation. The following inference is apparently legitimized by (P): from "He was happy that *p*" to *"He was happy a proposition" (cf. Asher (1987)). This is an objection from grammaticality. Horwich, concerning the more debated issue whether "that *p*" is always intersubstitutable with "the proposition that *p*" without loss of grammaticality, has noted that there are uncontroversial cases of co-referring terms where substitution, however, yields ungrammatical sentences. For instance, if we substitute "Pavarotti" for "The greatest tenor" in "The Italian singer Pavarotti never sings Wagner", we get *"The Italian singer the greatest tenor never sings Wagner" (see Schiffer (forthcoming)). Likewise, one cannot straightforwardly apply an existential generalisation to this sentence, since this would yield *"The Italian singer someone never sings Wagner". So, (P), which

seems to legitimize such inferences may be saved by appeal to independent grammatical facts.

It seems one could formulate something analogous to Quine's Thesis to shed further light on these cases. The thesis I have in mind is that one can infer a sentence "*F*(a proposition)" from "*F*(that *p*)" only if the expression "The thing such that *F*(it)" is well-formed. Clearly, one can infer "He believes a proposition" from "He believes that snow is white", and, indeed, the expression "The thing such that he believes it" is well-formed. However, concerning "He was happy that *p*", the corresponding description, *"The thing such that he was happy it" is not well-formed. These facts seem related, and indicate that there is something special about those expressions that yield sentences with "that"-clauses, but which are recalcitrant to (P). One can of course take this to show simply that "that"-clauses are not singular terms, but the obvious term-like behaviour in other cases, in particular, the obvious *predicate*-like behaviour of expressions that connect with "that"-clauses to form sentences, like "is true", "He believes", seems to me to show that an alternative solution would accommodate the data best. In fact, the very charge we have been discussing can easily be turned against those who deny that "that"-clauses are singular terms. The charge will then be that if expressions like "He was happy" are not to be considered as syntactic exceptions, why is it that it cannot, as opposed to other expressions which yield sentences with "that"-clauses, be formed into a description *"The thing such that he was happy it"? I believe these cases might require pragmatic treatment. For instance, a sentence "He was happy (surprised/upset/...) that *p*" might be elliptic for "He was happy (etc.) *to find out* that *p*", or some such. On this view, the expressions which do not satisfy the condition of well-formedness, as stated in the thesis, are indeed not predicates, but they are ellipses of predicates. This makes the claim that "that"-clauses are singular terms consistent with the denial that "He was happy" is a predicate, which, in turn, accounts for the apparent counter-examples to (P).

Finally, there is the inference from "He said something nice, namely, that *p*" to "It is nice that *p*" (Moltmann (2003: 89)). The latter should be understood as a grammatical variant of "That *p* is nice", as in 4.3. On the most natural interpretations of these sentences, however, they seem logically independent. In the first sentence, "is nice" does not seem to be predicated of the proposition. If it were, the inference should be unproblematic. It seems rather predicated of the act of saying. The sentence might, then, be equivalent to "He said something and his act of saying it was nice". A better analysis, it

seems to me, makes it equivalent to "He said something which it is nice to say" (i.e., "He said something such that to say *it* is nice"). Thus, the proper object of "is nice" here is "to say *x*", whereas in the latter sentence, its object is a "that"-clause. It is thus an equivocation of "nice". We need not worry here about how to categorize expressions of the form "to φ". Perhaps they refer to action-types. What is important is that we have, again, an ellipsis which seems to explain why the apparently invalid inference is indeed invalid.

5.7 ON ASSERTIBILITY

Much of what has been said in this book presupposes the correctness and feasibility of a non-traditional semantics, that is, a semantics where truth does not play any role. I will therefore give a few, admittedly speculative and freewheeling, comments on this aspect of the deflationist's dialectic situation. Logical space of course offers an unsurveyable amount of candidates for non-alethic semantics, but one normally thinks of the alternatives as "use-theoretic" accounts, i.e., roughly speaking, semantic theories taking as central concepts defined in terms of linguistic behaviour. It is natural to think of in-ferentialism as an exception here, but in my opinion, a correct inference is just one which preserves the central semantic property. Those expressions whose meaning is given an inferentialist treatment are simply those that are treated intralinguistically, but this leaves it open what relation between expressions or sentences that are relevant in accounting for the expression's meaning.

That deflationism has seemed to require such an alternative semantics has probably been the major reason for not endorsing it. On the other hand, though much technical work has been done within the Tarski-Davidson para-digm, much with the appearance of empirically well-evidenced, applied semantics, the debates about the very idea of thus explaining or illuminating meaning are inconclusive at best.[12] It simply remains unclear just what Tarskian truth-definitions have to do with meaning. One thing that remains fairly safe to say, however, is that truth-definitions for a set of sentences of a certain language needs to be coupled with much further theory in order that people's linguistic behaviour is touched upon. Use-theories make this con-

---

[12] See Soames (2003: 309ff.) on another discrepancy between philosophical discussions about Davidsonian semantics and its applications.

nection immediately. Here, then, is a sweeping and tentative (and biased) comparison between truth-theoretic and use-theoretic semantics: the former is simple and clean, and often gets things right (whether explanatory or not), while various use-theorists are lucky if they manage to say something that is both clear enough to judge and is not immediately subject to clear counter-examples. However, this disadvantage of use-theories is compensated by the relatively higher gains of success: namely an account which connects more with actual speakers and their linguistic behaviour, which ought to be a main *explanandum*.

The factual adequacy and simplicity of truth-theoretic descriptions are not necessarily traits that support them, further. A deflationist's view of this endeavour should be that it is a gathering of disguisedly folk-semantic de-scriptions and connections between them, which in the end do nothing by way of explanation. For instance, the truth-functional account of sentential connectives has been proved by Field (1994a: 256ff.) to be derivable from a simple disquotational truth-theory (another example of this type of trivial-isation was given in 2.4). Truth-theoretic semantics, from this perspective, is merely a matter of applying those concepts (and various complex derivative concepts) that a more thoroughgoing theory should take as its data. The more complex the truth-theoretic accounts, the more difficult it is for the defla-tionist to prove that this is the case, of course. But this is no reason to doubt that they can eventually be thus trivialised. In fact, no arguments that this should be impossible have, to my knowledge, been presented. Of course, many insights of, shall we say, a "structural" kind, may have been gained by truth-theoretic semantics along the way, and should be kept. But the basic substantial question of what central semantic concept to employ seems, on balance, quite open.

Let me in these closing pages express my own sympathies and interests pertaining to this basic question. The notion of assertibility is usually thought of as epistemological. However, this seems to make it conceptually dependent on propositional and intentional notions, to wit, on the notion of belief in a proposition. Besides ruled out by (NC), assertibility thus understood will be unfit for playing a role in semantics for the same reason that propositional truth is: it presupposes that one know what it is for a sentence to express a proposition, but this is too close to that which the semantics should explain (cf. the argument in 4.4). To explain the meaning of an expression or sentence by recourse to the assertibility (conditions) of sentences thus requires that assertibility is primarily a property of sentences, not of propositions.

Accordingly, I believe semantics should be entirely naturalistic in the sense that intentional terms are ruled out from the meta-language. But note that this naturalism will not be of the demanding kind required by a truth-theoretic semantics. For giving naturalistic *truth*-conditions of sentences with intentional vocabulary requires naturalistic *reduction*, this being the nature of truth-conditions. A use-theoretic naturalist semantics, by contrast, need only avail itself of defeasible assertibility conditions of sentences with intentional terms, plausibly taking such conditions to concern people's observable behaviour and environment. This is a type of "interpretivist" semantics, but not one which states *equivalences* between belief-sentences, etc. and sentences describing behaviour (or, even worse, observers' attributions of beliefs, etc.). This, I think, is more in line with the intuitive idea that one can always be wrong about what people believe, even if one is right about how they behave, and use this knowledge appropriately for deriving belief-attributions. If a correct statement of such defeasible assertibility conditions can also be coupled with an explanation of the function of expressions governed by such assertibility conditions, then it can also be argued that one has given an explanation of the cognitive state itself, though, of course, in the same type of indirect way that (D) is taken as an explanation of truth itself.

We have not yet said what assertibility is, only that (NC) rules out any definition in intentional terms. Possible candidates, then, include that of an utterance simply being assented to in certain circumstances. More sophisticated are the notions of correctness and incorrectness relative to a community, explained in terms of positive and negative reinforcements, which can themselves be naturalistically explained. Thus, on Brandom's development of Haugeland's (1982) account of norms, "[p]ositively reinforcing a disposition to produce a performance of a certain kind as a response to a stimulus of a certain kind is responding to the response in such a way as to make it more likely in the future that a response of that kind will be elicited by a stimulus of the corresponding kind" (1994: 35). Negative reinforcement is similarly defined as a response making it less likely that the subject's behaviour will be repeated in the given kind of circumstance. A sentence can then be said to have *C* as its necessary and sufficient assertibility conditions iff utterances of the sentence tend to be positively reinforced by the community members when *C* obtain and negatively reinforced when *C* do not obtain. Being subject to positive reinforcement may look as an intuitive account of some notion of "correct", which is still entirely naturalistic. (On the other hand, one may discuss whether the semantic concept is more reasonably taken to be the final

state of competent speakers, i.e., their speech dispositions, rather than what looks more like the causal route to this state, i.e., the reinforcements.)

A different use-theory takes into account not only the conditions under which sentences are (correctly) uttered, but also the consequences of uttering them, thus giving more of a "causal role semantics". Such an idea may be motivated by the heuristic principle that we should look at simpler examples of the phenomenon to be explained (language), and then extend the theory for these to the more complex phenomenon. To account for the signals of simpler animals, it seems sufficient to describe precisely the causal role of productions of signal-types. A typical cause could be, e.g., the presence of a predator, and the typical effect a fugitive behaviour of conspecifics. Why could not human languages be simply immensely more complex systems of sounds with associated causal roles? Thanks to superior intelligence and memory, the causes and effects of utterances can be much more distant and indirect than those of the signals of simpler animals. Further, human languages are compositional, so subsentential expressions will contribute systematically to the causal roles of sentences where they occur. On this type of theory, then, the central semantic property of a primitive expression is its causal role, which, in turn, is identical to the contribution it makes to the causal roles of sentences containing it. Expressions that intuitively lack empirical content will be treated intralinguistically, and will serve the purpose of partaking in sentences that function as bridges between sentences which do have empirical content. Insofar as a sentence has assertibility conditions that eventually depend, if only *via* other sentences, on environmental features, they have a causal role of practical relevance.

Of course, on all these theories, there will be non-semantic inhibiting factors complicating the story. One does not, for instance, normally make irrelevant utterances, even when their assertibility conditions obtain. Otherwise, we would constantly utter analytic sentences, since their assertibility conditions always obtain. But as long as the inhibiting factors can be isolated fairly well, the semantic properties of expressions can be isolated as well. For any linguistic piece of behaviour, there is intuitively a semantic part of the explanation to the behaviour, and all kinds of non-semantic ones. There are cognitive factors (different speakers will assent to different sentences because they have undergone different perceptual stimulations, etc.), social (speakers may differ in humour), and so on and so forth. Irrelevance, by the way, seems to be a notion particularly suitable for a causal explanation (cf. Sperber and Wilson (1986: Ch. 3)).

I would also like to note that there is an important sense in which we may still appeal to the truth-conditions of certain sentences when accounting for their meaning. We may do so in the special sense that we may take the assertibility-conditions of those sentences to be those in which, as we casually have it, they are true. For instance, we are free to say, e.g., that the semantic meaning of "It is raining" consists in the fact that speakers are conditioned to utter it when and only when it is raining (at the place of utterance). Here, we have not *used* "true", but only given a condition for assertibility (not defined in terms of truth), which happens to be the condition under which, casually speaking, it says something true. Of course, attitude-ascribing sentences will not be explained on this model. Stating that "A believes that *p*" is assertible when and only when A believes so and so (roughly speaking) will of course violate (NC).

There is good reason to take simple empirical sentences to have precisely such assertibility conditions. Consider the predator-indicating signal of a simple animal. It may first be thought that the typical cause of this signal should be taken to be a certain kind of perception, perhaps referred to as a certain series of stimulations on the retina. But it may happen that the predator always makes a certain noise, which the signalling animal comes to associate with the visual stimuli in such a way that it will emit the signal upon hearing the noise, even absent the visual stimuli. We should not say then that the signal has changed in meaning. It was always a sound indicating the presence of a predator, independently of what causal chains between the predator and the signal. Therefore, although the signalling animal will emit the signal more often in conditions individuated in terms of stimuli than in conditions in which there is a nearby predator, the latter makes more sense to take as semantically relevant.

Plausibly, the causal roles of a type of sound-pattern should be such that the cause is what motivates the effect. The fugitive behaviour should not ultimately be explained by reference to the perceptual stimulation of the conspecific, but to what tends to cause those stimulations, namely, the presence of a predator, since the latter is what makes the effect of the signal beneficial. Language, on this model, is at base an extension of the senses. Instead of having to be there to see or hear what is happening, one takes the requisite measures thanks to other individuals' seeing or hearing it. With greater intelligence, the language can develop so as to enable sound-patterns with more intricate connections to observation-related sound-patterns, thus

creating sentential connectives, quantifiers, non-referring singular terms that can instantiate quantifications, etc.

A word, of course, has an infinite amount of causal roles of various kinds. The task of the semanticist is to capture that role which is semantically relevant. Intuitively, the meaning of an unambiguous word is a single thing that partakes in the semantic part of the explanation of any piece of linguistic behaviour involving the word, and does so uniformly. On this account, this "thing" will be a general pattern, namely, a general contribution to the causal roles of sentences. These roles will mesh in complicated ways with non-semantic factors, such as perceptions and inferences. Sometimes, people end up saying the wrong thing, like "The Earth is flat". This may both be universally assented to and even positively reinforced.[13] This does not contradict the idea of a dispositionalist or assertibility-conditional semantics. For each word in this sentence has a general effect on sentences, and the most basic sentences containing these words will always and only be assented to in the right circumstances. These are sentences such that the things and properties figuring essentially in the specification of the causal roles of their words are readily perceptually accessible. It is plausibly by exposure to such basic sentences that the words are learnt. The semantically relevant causal roles of words (i.e., their contribution to the roles of *any* sentence containing them) should be sought by looking at their roles in these basic sentences. The reason "The Earth is flat" is uttered has as its semantic explanation the relevant causal roles of the words therein, and, as its non-semantic explanation, the fact that certain perceptions and non-demonstrative inferences have been made. The semantic content of the whole sentence is just given by giving the general causal roles of its parts and its mode of composition. Taking this as input, together with (very many) non-semantic facts, including assumptions about cognitive normal functioning, a speakers' behaviour will follow as output. In the present case, this will be the utterance of "The Earth is flat". The same causal roles will determine the effect that an utterance has on other speakers, given their general cognitive state.

---

[13] Though this is not the main response to the alleged problem, I believe that the possible positive reinforcement of utterances of "The Earth is flat" will be qualitatively different from those that institute semantic competence. When teaching a child how to speak, different sanctions are applied than when teaching social codes or extralinguistic knowledge. Mere imitation and sensitivity to "not being understood", i.e., not prompting appropriate conversational reactions may be powerful cues in acquiring semantic competence.

An explanation in the same spirit should be given for cases in which we fail to assent to an analytic sentence, perhaps because it is very complex. Now, how can one say that the assertibility conditions for a sentence are unconditionally fulfilled although perhaps all speakers fail to assent to it? Obviously, the assertibility conditions of a sentence will not be such that always, when an utterance is made in that condition, there will be assent and a positive reinforcement. A plausible assignment of an assertibility condition C for a sentence s is had by balancing the actual reinforcement conditions with simplicity. Thus, a good assignment should be both such that utterances of the sentence are normally positively sanctioned in those conditions and negatively sanctioned outside, and such that learners of the language can readily internalise the condition. So, even if there were conditions one could formulate for the assertibility of conditionals which would include those cases where speakers fallaciously assent to conditionals, we ought to formulate them in a simpler way and accept the slack between the formulation and speakers' sanctions. We will then get general, non-universal assertibility conditions of sentences as their central semantic content, statements of the form "In general, s is assertible iff p". And now, we can simply say that analytic sentences are those whose general assertibility conditions are trivially fulfilled. Suppose we take the central semantic character of conditionals to be the fact that, in general, a conditional is assertible iff the consequent is assertible if the antecedent is. And suppose we take the semantic character of disjunctions to be that, in general, a disjunction is assertible iff either disjunct is assertible if the negation of the other is assertible. Now, we can see that the general assertibility conditions of the sentence "If it is raining or it is snowing, then if it is not snowing, it is raining" are trivially fulfilled. If speakers do not always assent to this, this just shows that various cognitive factors prevent their semantic competence from being properly executed. Something along these lines could perhaps even be used to argue for the analyticity of mathematics.

Finally, we may address the question what motivates the claim that the above proposals are candidates for *semantic* theories, i.e., that they could somehow explain *meaning*, given that this notion is ruled out from the language of the theory? Since we do not reduce this term naturalistically, there must be some other connection between it and the claims of the semantics. But note that the semantic theory, if complete, will also describe the semantic character of the words "mean" and "meaning", in a systematic way. The defence of the claim that the theory is worthy of the label "sem-

antic" must eventually connect somehow with this account. What must be shown is that the naturalistic properties of sentences that the theory takes as essentially semantic are also those that are *tracked* when we use "mean" and "meaning" (or, better, "literal meaning"). What this tracking comes to exactly may be a matter of discussion. But since the assertibility conditions of sentences containing intentional vocabulary will be ones concerning speakers' behavioural dispositions in relation to events in the environment, and since it is such facts that are taken as essentially semantic by the semantic theory, it might seem that there is some hope of explaining what the tracking comes in such a way that labelling the theory as "semantic" is justified.

These were some very programmatic and preliminary ideas about how to replace a Platonist, representational outlook with a non-reductive naturalistic one. Despite being at such an embryonic stage, it should be clear how direct the motivational connection with the preceding claims. Deflationism and nominalism separately rule out the use of truth in the metalanguage, and propositional attitudes are ruled out from the metalanguage because of their association with propositions. These notions are to be treated linguistically by the naturalistic, non-representational semantics, but, hopefully, still in such a way as to justify the claim that they have been exhaustively explained.

References

ARMOUR-GARB, B. P. and J. C. BEALL (eds.) (2005), *Deflationary Truth*, Chicago and La Salle, Ill., Open Court.

ASHER, N. (1987), "A Typology of Attitude Verbs and Their Anaphoric Properties", *Linguistics and Philosophy* **10**, pp. 127-89.

AUSTIN, J. L. (1950), "Truth", *Proceedings of the Aristotelian Society*, Supp. Vol. XXIV, pp. 111-28.

AYER, A. J. (1935), "The Criterion of Truth", *Analysis* **3**, pp. 28-32.

—— (1963), *The Concept of a Person and Other Essays*, London, Macmillan.

BAKER, G. P. and P. M. S. HACKER (1980), *Wittgenstein: Understanding and Meaning. Volume 1 of An Analytical Commentary on the Philosophical Investigations*, Oxford, Basil Blackwell.

BALDWIN, T. (1989), "Can There Be a Substantive Theory of Truth?", *Recherche sur la Philosophie et le Language* **10**, pp. 99-118.

BAR-ON, D., C. HORISK and W. LYCAN (2000), "Deflationism, Meaning and Truth-conditions", *Philosophical Studies* **101**, pp. 1-28.

BEALER, G. (1982), *Quality and Concept*, Oxford, Oxford University Press.

—— (1998), "Propositions", *Mind* **107**, pp. 1-32.

BEALER, G. and U. MÖNNICH (1989), "Property Theories", in D. Gabbay & F. Guenthner (eds.), *Handbook of Philosophical Logic*, Vol. IV, Dordrecht, D. Reidel.

BEEBE, J. R. (2003), "Attributive Uses of Prosentences", *Ratio* **16** (new series), pp. 1-15.

BLACK, M. (1948), "The Semantic Definition of Truth", *Analysis* **8**, pp. 49-63.

BLACKBURN, S. (1984), *Spreading the Word*, Oxford, Clarendon Press.

BLACKBURN, S. and K. SIMMONS (eds.) (1999), *Truth*, New York, Oxford University Press.

BOGHOSSIAN, P. (1990), "The Status of Content", *The Philosophical Review* **99**, pp. 157-84.

BRANDOM, R. (1994), *Making It Explicit*, Cambridge, MA, Harvard University Press.

—— (1997), "Replies to Commentators", in Villanueva (1997), pp. 199-214.

—— (2000), "Vocabularies of Pragmatism: Synthesizing Naturalism and Historicism", in R. Brandom (ed.), *Rorty: and His Critics*, Oxford Basil Blackwell, pp. 156-82.

—— (2002), "Explanatory vs. Expressive Deflationism about Truth", in Schantz (2002), pp. 103-19.

BRENTANO, F. C. (1904), *The True and the Evident*, ed. and trans. by R. M. Chisholm, London, Routledge and Kegan Paul, 1966.

BUCHANAN, R. (2003), "Are Truth and Reference Quasi-Disquotational?", *Philosophical Studies* **113**, pp. 43-75.

BURGESS, J. P. (2002), "Is There a Problem about the Deflationary Theory of Truth?", Halbach, V and L. Horsten (2002), pp. 37-55.

CARNAP, R. (1942), *Introduction to Semantics and Formalisation of Logic*, Cambridge, MA., Harvard University Press.

—— (1956), *Meaning and Necessity*, Chicago, University of Chicago Press.

CHAPUIS, A. and A. GUPTA (eds.) (2000), *Circularity, Definition and Truth*, New Delhi, Indian Council of Philosophical Research.

CHIHARA, C. (1979), "The Semantic Paradoxes: A Diagnostic Investigation", *The Philosophical Review* **88**, pp. 590-618.

—— (1984), "The Semantic Paradoxes: Some Second Thoughts", *Philosophical Studies* **45**, pp. 223-9.

CHISHOLM, R. (1977), *Theory of Knowledge* (second edition), Englewood Cliffs, Prentice-Hall.

COHEN, J. (1950), "Mr. Strawson's Analysis of Truth", *Analysis* **10**, pp. 136-40.

COZZO, C. (1994), *Meaning and Argument*, Stockholm Studies in Philosophy, no. 17, Stockholm, Almqvist & Wiksell International.

DAVID, M. (1994), *Correspondence and Disquotation*, Oxford, Oxford University Press.

DAVIDSON, D. (1967), "Truth and Meaning", *Synthèse* **17**, pp. 304-23, references are to Davidson (1984).

—— (1984), *Inquiries into Truth and Interpretation*, Oxford, Clarendon Press.

—— (1986), "A Coherence Theory of Truth and Knowledge", in LePore (1986).

—— (1990), "The Structure and Content of Truth", *The Journal of Philosophy* **87**, pp. 282-309.

—— (1994), "What is Quine's View of Truth?", *Inquiry* **37**, pp. 437-40.

—— (1996), "The Folly of Trying to Define Truth", *The Journal of Philosophy* **93**, pp. 263-78.

DEVITT, M. (1984), *Realism and Truth*, Oxford, Basil Blackwell.

—— (2002), "The Metaphysics of Deflationary Truth", in Schantz (ed.) (2002), pp. 60-78.

DEWEY, J. (1938), *Logic: The Theory of Inquiry*, New York, Holt, Rinehart & Winston.

DODD, J. (2000), *An Identity Theory of Truth*, New York, St. Martin's Press.

DUMMETT, M. (1959), "Truth", in Dummett (1978), pp. 1-19.

—— (1963), "The Philosophical Significance of Gödel's Theorem", *Ratio* **5**, pp. 140-55, references are to Dummett (1978).

—— (1973), *Frege: The Philosophy of Language, Second Edition*, Cambridge, MA., Harvard University Press.

—— (1978), *Truth and Other Enigmas*, Cambridge, MA., Harvard University Press.

—— (2003), "Truth and the Past, Lecture 1: The Concept of Truth", *Journal of Philosophy* **100**, pp. 5-25.

ETCHEMENDY, J. (1988), "Tarski on Truth and Logical Consequence", *Journal of Symbolic Logic* **53**, pp. 51-79.

EVANS, G. and J. MCDOWELL (eds.) (1976), *Truth and Meaning*, Oxford, Clarendon Press.

FIELD, H. (1972), "Tarski's Theory of Truth", *The Journal of Philosophy* **69**, pp. 347-75.

—— (1980), *Science without Numbers*, Princeton, Princeton University Press.

—— (1984), "Is mathematical knowledge just logical knowledge?", *The Philosophical Review* **93**, pp. 509-52.

—— (1986), "The Deflationary Conception of Truth", in MacDonald and Wright (1986).

—— (1989), *Realism, Mathematics, and Modality*, Oxford, Basil Blackwell.

—— (1992), "Critical Notice: Paul Horwich's *Truth*", *Philosophy of Science* **59**, pp. 321-30.

—— (1994a), "Deflationist Views about Meaning and Content", *Mind* **103**, pp. 249-85.

—— (1994b), "Disquotational Truth and Factually Defective Discourse", *The Philosophical Review* **103**, pp. 405-52.

—— (2001), *Truth and the Absence of Fact*, Oxford, Clarendon Press.

<div align="center">References</div>

FORBES, G. (1986), "Truth, Correspondence and Redundancy", in MacDonald and Wright (1986), pp. 27-54.

FREDDOSO, A. J., and H. SCHUURMAN (1980), *Ockham's Theory of Propositions: Part II of the Summa Logicae* (trans.), Notre Dame, Ind., University of Notre Dame Press.

FREGE, G. (1884/1950), *The Foundations of Arithmetic*, trans. by J. L. Austin, Cambridge, Basil Blackwell.

—— (1892/1970), "On Sense and Reference", in Geach and Black (1970), pp. 56-78.

—— (1918/1984), "Thoughts", trans. by P. Geach and R. H. Stoothoff in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness (1984), Oxford, Basil Blackwell.

GARCÍA-CARPINTERO, M. (1998), "A Paradox of Truth Minimalism", in Martínez, Rivas and Villegas (1998), pp. 37-63.

GEACH, P. (1960), "Ascriptivism", *Philosophical Review* **69**, pp. 221-5.

—— (1967), "Intentional Identity", *Journal of Philosophy* **64**, pp. 627-32.

—— (1975), "Names and Identity", in S. Guttenplan (ed.), *Mind and Language: Wolfson College Lectures* 1974, Oxford, Clarendon Press.

GEACH, P. and M. BLACK (eds.) (1970), *Translation from the Philosophical Writings of Gottlob Frege*, Oxford, Basil Blackwell.

GROVER, D., CAMP, J. and BELNAP, N. (1975), "A Prosentential Theory of Truth", *Philosophical Studies* **27**, pp. 73-125.

GUNDERSON, K. (ed.) (1981), *Language, Mind and Knowledge*, Minneapolis, University of Minnesota Press.

GUPTA, A. (1993a), "Minimalism", *Philosophical Perspectives* **7**: *Language and Logic*, pp. 359-69.

—— (1993b), "A Critique of Deflationism", *Philosophical Topics* **21**, pp. 57-81.

HAACK, S. (1978), *Philosophy of Logics*, Cambridge, Cambridge University Press.

HALBACH, V. (2001), *Semantics and Deflationism*, unpublished habilitation thesis.

—— (2002), "Modalized Disquotationalism", in Halbach and Horsten (2002), pp. 75-102.

HALBACH, V. and L. HORSTEN (eds.) (2002), *Principles of Truth*, Frankfurt a. M., Hänsel-Hohenhausen.

HALE, B. (1987), *Abstract Objects*, Oxford, Basil Blackwell.

References

HALE, B. and C. WRIGHT (eds.) (1997), *A Companion to the Philosophy of Language*, Cambridge, Basil Blackwell.

—— (2001), *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*, Oxford, Oxford University Press.

HARMAN, G. and D. DAVIDSON (eds.) (1972), *Semantics of Natural Language*, Dordrecht, D. Reidel.

HAUGELAND, J. (1982), "Heidegger on Being a Person", *Noûs* **16**, pp. 15-26.

HECK, R. (unpublished), "What is a Singular Term?", unpublished manuscript.

HEMPEL, C. (1935a), "On the Logical Positivists' Theory of Truth", *Analysis* **2**, pp. 49-59.

—— (1935b), "Some Remarks on 'Facts' and Propositions", *Analysis* **2**, pp. 93-6.

HILL, C. (2002), *Thought and World: An Austere Portrayal of Truth, Reference, and Semantic Correspondence*, Cambridge, Cambridge University Press.

HORWICH, P. (1998a), *Truth* (second edition), Oxford, Clarendon Press.

—— (1998b), *Meaning*, Oxford, Clarendon Press.

—— (2002), "A Defence of Minimalism", in Halbach and Horsten (2002), pp. 57-73.

JAMES, W. (1907), "Pragmatism's Conception of Truth", reprinted in Lynch (2001), pp. 211-28. Page references are to the latter work.

JOHNSON, W. E. (1921), *Logic*, Cambridge, Cambridge University Press.

KALDERON, M. E. (1997), "The Transparency of Truth", *Mind* **106**, pp. 475-98.

KATZ. J. (1998), *Realistic Rationalism*, Cambridge, MA., MIT Press.

KING, J. (2002), "Designating Propositions", *Philosophical Review* **111**, pp. 341-71.

KIRKHAM, R. (1992), *Theories of Truth*, Cambridge, MA., MIT Press.

KOVACH, A. (1997), "Deflationism and the Derivation Game", *Mind* **106**, pp. 575-9.

KRIPKE, S. (1976), "Is There a Problem about Substitutional Quantification?" in Evans and McDowell (1976), pp. 325-419.

LANCE, M. (1996), "Quantification, Substitution, and Conceptual Content", *Noûs* **30**, pp. 481-507.

—— (1997), "The Significance of Anaphoric Theories of Truth and Reference", in Villanueva (1997), pp.181-98.

LEEDS, S. (1978), "Theories of Reference and Truth", *Erkenntnis* **13**, pp. 111-30.

LEPORE, E. (ed.) (1986), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York, Blackwell.

LEWIS, D. (1972), "General Semantics", in Harman and Davidson (1972).

—— (1981), "Languages and Language" in Gunderson (1981).

LEWY, C. (1947), "Truth and Significance", *Analysis* **8**, pp. 24-7.

LYNCH, M. (2001), *The Nature of Truth: Classic and Contemporary Perspectives*, Cambridge, MA., MIT Press.

MACDONALD, G. and C. WRIGHT (eds.) (1986), *Fact, Science and Morality*, Oxford, Basil Blackwell.

MACKIE, J. L. (1973), *Truth, Probability, and Paradox*, Oxford, Oxford University Press.

MARTIN-LÖF, P. (1991), "A Path from Logic to Metaphysics", in Costantini, D. and Galavotti, M.C. (eds.) (1991), *Nuovi problemi della logica e della filosofia della scienza*, vol. I, Bologna, Clueb, pp. 141-9.

MARTÍNEZ, C., U. RIVAS AND L. VILLEGAS (1998), *Truth in Perspective*, Ashgate, Aldershot.

MATTHEWS, F. (1994), "The Measure of Mind", *Mind* **103**, pp. 131-46.

MCGEE, V. (2000), "The Analysis of '*x* is true' as 'For any *p*, if *x*="*p*", then *p*'", in Chapuis and Gupta (2000).

—— (2005), "A Semantic Conception of Truth?", in Armour-Garb and Beall (2005), pp. 111-42.

MCGINN, C. (2002), "The Truth about Truth" in Schantz (2002), pp. 194-204.

MCGRATH, M. (1997a), "Weak Deflationism", *Mind* **106**, pp. 69-98.

—— (1997b), "Reply to Kovach", *Mind* **106**, pp. 581-6.

—— (2003), "What the Deflationist May Say about Truthmaking", *Philosophy and Phenomenological Research* **66**, pp. 666-88.

MOLTMANN, F. (2003), "Propositional Attitudes without Propositions", *Synthese* **135**, pp. 77-118.

NÆSS, A. (1938), " 'Truth' as Conceived by Those Who Are Not Professional Philosophers", in *Skrifter utgitt av Det Norske Videnskaps-Akademi I Oslo II. Hist.- Filos. Klasse*, vol. IV, Oslo, I kommisjon hos J. Dybwad.

NEURATH, O. (1934), "Radical Physicalism and the Real World", in *Philosophical Papers: 1913-1946*, (1983), eds. R. Cohen and M. Neurath, Dordrecht, D. Reidel.

PAGIN, P. (2001), "Frege on Truth and Judgment", *Organon F* **8**, pp. 1-13.

References

PAP, A. (1952), "Note on the 'Semantic' and 'Absolute' Concept of Truth", *Philosophical Studies* **3**, pp.1-8.

PARSONS, T. (1993), "On Denoting Propositions and Facts", *Philosophical Perspectives* **7**, pp. 441-60.

PEIRCE, C. S. (1878), "How To Make Our Ideas Clear", reprinted in Lynch (2001), pp. 193-209. Page references are to the latter work.

PITCHER, G. (ed.) (1964), *Truth*, Englewood Cliffs, NJ, Prentice Hall.

PLATTS, M. (1979), *Ways of Meaning*, London, Routledge and Kegan Paul.

PRAWITZ, D. (1998), "Truth from a Constructive Perspective", in Martínez, Rivas and Villegas (1998), pp. 23-35.

PRICE, H. (1988), *Facts and the Function of Truth*, Oxford, Basil Blackwell.

PRIOR, A. N. (1971), *Objects of Thought*, Oxford, Clarendon Press.

PUTNAM, H. (1981), *Reason, Truth, and History*, Cambridge, Cambridge University Press.

QUINE, W. V. (1951), "Two Dogmas of Empiricism", in Quine (1953), pp. 20-46.

—— (1953), *From a Logical Point of View*, Cambridge, MA, Harvard University Press.

—— (1960), *Word and Object*, Cambridge, MA., MIT Press.

—— (1970), *Philosophy of Logic*, Englewood Cliffs, NJ, Prentice Hall.

—— (1974), *The Roots of Reference*, La Salle, Ill., Open Court.

—— (1981), *Theories and Things*, Cambridge, MA., Harvard University Press.

—— (1987), *Quiddities*, Cambridge, MA., Harvard University Press.

—— (1990), *Pursuit of Truth*, Cambridge, MA, Harvard University Press.

—— (1991), "Two Dogmas in Retrospect", *Canadian Journal of Philosophy* **21**, pp. 265-74.

RAATIKAINEN, P. (2005), "Horwich's Way Out", *Analysis* **65**, pp. 175-7.

RAMSEY, F. (1927), "Facts and Propositions", *Proceedings of the Aristotelian Society*, supp. vol. 7, pp.153-70.

RESNIK, M. D. (1990), "Immanent Truth", *Mind* **99**, pp. 405-24.

—— (1997), *Mathematics as a Science of Patterns*, Oxford, Clarendon Press.

RUSSELL, B. (1910), "On the Nature of Truth and Falsehood", in *Logical and Philosophical Papers Vol.6 1909-13* (1966), New York, Simon and Schuster, pp. 116-24.

—— (1912), *The Problems of Philosophy*, Oxford, Oxford University Press.

—— (1918), *The Philosophy of Logical Atomism*, D. Pears (ed.), Open Court, La Salle, Ill.

References

SCHANTZ, R. (ed.) (2002), *What is Truth?*, Berlin, de Gruyter.

SCHIFFER, S. (1987), *The Remnants of Meaning*, Cambridge, MA, MIT Press

—— (1992), "Belief Ascription", *Journal of Philosophy* **92**, pp. 499-521.

—— (1996), "Language-Created, Language-Independent Entities", *Philosophical Topics* **24**, pp. 149-67.

—— (forthcoming), "Propositional Content", in E. Lepore and B. Smith, *Oxford Handbook of Philosophy of Language*.

SCHLICK, M. (1935), "Facts and Propositions", *Analysis* **2**, pp. 65-70.

SHAPIRO, S. (1998), "Proof and Truth: Through Thick and Thin", *Journal of Philosophy* **95**, pp. 493-521.

SMITH, P. (1998), "Approximate Truth for minimalists", *Philosophical Papers* **27**, pp. 119-28.

SOAMES, S. (1999), *Understanding Truth*, Oxford, Oxford University Press.

—— (2003), *Philosophical Analysis in the Twentieth Century, Vol. 2, The Age of Meaning*, Princeton, Princeton University Press.

SOSA, E. (1993), "The Truth of Modest Realism", *Philosophical Issues* 3, *Science and Knowledge*, pp. 77-95.

SPERBER, D. and D. WILSON (1986), *Relevance: Communication and Cognition*, Cambridge, MA., Harvard University Press.

SRZEDNICKI, J. (1966), "It Is True", *Mind* **75**, pp. 385-95.

STRAWSON, P. F. (1949), "Truth", *Analysis* **9**, pp. 83-97.

—— (1950), "Truth", *Proceedings of the Aristotelian Society*, supp. vol. 24, pp. 129-56. (All references are to Pitcher (1964).)

—— (1952), *Introduction to Logical Theory*, London, Methuen.

—— (1964), "A Problem about Truth – Reply to Warnock", in Pitcher (1964), pp. 68-84.

—— (1974), *Freedom and Resentment and Other Essays*, London, Methuen.

TARSKI, A. (1944), "The Semantic Conception of Truth", *Philosophy and Phenomenological Research* **4**, pp. 341-76.

TOMBERLIN, J. E. (1997), "Quantification: Objectual or Substitutional?", in Villanueva (1997), pp. 155-67.

TYE, M. (1989), *The Metaphysics of Mind*, Cambridge, Cambridge University Press.

VAN INWAGEN, P. (2002), "Generalizations of Homophonic Truth-sentences", in Schantz (2002), pp. 205-22.

VILLANUEVA, E., (ed.) (1997), *Truth* (*Philosophical Issues* **8**), Atascadero, CA., Ridgeview Publishing Company.

References

VISION, G. (1997), "Why Correspondence Truth Will Not Go Away", *Notre Dame Journal of Formal Logic* **38**, pp. 104-31.

WARNOCK, G. J. (1964), "A Problem about Truth", Pitcher (1964) pp. 54-67.

WILLIAMS, C. J. F. (1976), *What Is Truth?*, Cambridge, Cambridge University Press.

WILLIAMS, M. (1977), *Groundless Belief*, New Haven, Yale University Press.

—— (1986), "Do We (Epistemologists) need a Theory of Truth?", *Philosophical Topics* **14**, pp. 223-42.

WITTGENSTEIN, L. (1922), *Tractatus Logico-Philosophicus*, London, Routledge and Kegan Paul.

—— (1953), *The Philosophical Investigations*, trans. by G. E. M. Anscombe, Oxford, Basil Blackwell.

—— (1961), *Notebooks 1914-1916*, Oxford, Basil Blackwell.

—— (1978), *Remarks on the Foundations of Mathematics*, eds. G. H. von Wright, R. Rhees and G. E. M. Anscombe, trans. by G. E. M. Anscombe, Oxford, Basil Blackwell.

WRIGHT, C. (1983), *Frege's Conception of Numbers as Objects*, Aberdeen, Aberdeen University Press.

—— (1992), *Truth and Objectivity,* Cambridge, MA, Harvard University Press.

—— (1999), "Truth: A Traditional Debate Reviewed", in Blackburn and Simmons (1999), pp. 203-38.

YNGVE, V. H. (1960), "A Model and an Hypothesis for Language Structure", *Proceedings of the American Philosophical Society* **104**, pp. 444-66.

INDEX