To be published in *Synthèse*. The original publication is available at www.springerlink.com, DOI 10.1007/s11229-012-0163-2.

# FORMULATING DEFLATIONISM

**Arvid Båve, Gothenburg University**

ABSTRACT: I here argue for a particular formulation of truth-deflationism, namely, the propositionally quantified formula, (Q) "For all *p*, <*p*> is true iff *p*". The main argument consists of an enumeration of the other (five) possible formulations and criticisms thereof. Notably, Horwich's Minimal Theory is found objectionable in that it cannot be accepted by finite beings. Other formulations err in not providing non-question-begging, sufficiently direct derivations of the T-schema instances. I end by defending (Q) against various objections. In particular, I argue that certain circularity charges rest on mistaken assumptions about logic that lead to Carroll's regress. I show how the propositional quantifier can be seen as on a par with first-order quantifiers and so equally acceptable to use. While the proposed parallelism between these quantifiers is controversial in general, deflationists have special reasons to affirm it. I further argue that the main three types of approach the truth-paradoxes are open to an adherent of (Q), and that the derivation of general facts about truth can be explained on its basis.

## *I Deflationism and its formulation problem*

The *basic deflationist idea* (as I will call it) is that the following claims, or some generalisation covering them, is "all there is to say" about truth: that <snow is white> is true iff snow is white, that <grass is green> is true iff grass is green, and so on. (Instances of "<*p*>" refer to propositions, sentences, or some other type of truth-bearer.) The sentences expressing these claims are instances of the schema,

(T) <*p*> is true iff *p*.

I will call these instances *(T)-instances*, but I will use this expression sometimes to refer to propositions rather than sentences, depending on which is appropriate in the context. Nothing

of what I have to say until section 6, where I finally decide on a specific reading, depends on this choice. Now, a generalisation covering the (T)-instances, if there is one, would thus, according to the deflationist, be an *exhaustive* truth-theory, i.e., one which "does" everything a truth-theory should "do". A common idea, deriving from Paul Horwich (1998), is that a truth-theory is exhaustive just in case it explains, together with facts and laws that do not concern truth, all facts about truth (a theory of truth, however, is of course not one about the word "true" or the concept *true*, which require separate theories having different *explananda* and adequacy constraints). (What I called "the basic deflationist idea", however, I believe was first conceived by Frank Ramsey (1927) rather than Horwich).

The deflationary truth-theory, of whatever form, must plausibly entail the instances of (T) (or at least the non-pathological ones, but I will omit this qualification until we discuss the paradoxes separately in section 7). However, this constraint plausibly holds for non-deflationist truth-theories as well. A deflationary truth-theory must therefore be subject to further constraints. Intuitively, there should be a constraint (or desideratum) that the theory be a (relatively) *direct* generalisation over the instances. This is simply a reasonable interpretation of the basic deflationist idea, that the (T)-instances or a generalisation over them *suffice* as an exhaustive theory of truth (so that no more need or *should* be added). Inflationary truth-theories, by contrast, are clearly not subject to any such constraint, since their proponents gladly affirm that the nature of truth consists in its relation to other things, like facts, coherence of beliefs, etc. What could it mean that the theory should be a "direct" generalisation of the instances? I propose to spell this out by stating three subconstraints (to

be called "directness desiderata"), such that the general constraint of providing a direct generalisation of the (T)-instances is satisfied to the extent that the directness desiderata are. The first directness desideratum is that we should be able to derive the (T)-instances from the theory *relying only on logic*, and thus not on any richer metaphysical or conceptual assumptions. If there is no principled demarcation of logic, but only a continuum of more or less logical expressions, the desideratum is rather that the inferences relied on should be *relatively* logical, i.e., toward the logical end of the continuum. The second directness desideratum is that we be able to derive the (T)-instances using as few inference-steps as possible (that this is a *directness* desideratum is obvious enough). Thirdly, as little as possible *other than* the instances should be entailed by the theory. This is in clear contradistinction to inflationary theories which involve many other concepts (e.g., of fact, correspondence, coherence, success, etc.) and which will entail many (non-tautological) claims involving those concepts. It also captures the intent of the phrase "being a generalisation of". A "generalisation of", "$Fa_1$", …, "$Fa_n$" is of course "$\forall xFx$", rather than, say, "$\forall xFx\&Gx$", even if the latter entails "$Fa_1$", …, "$Fa_n$". These desiderata seem to be reasonable explications of the "basic deflationary idea" itself, and should be accepted by deflationists. Satisfying them also means staying true to Horwich's (1998: 11) idea that the concept of truth is *pure* in the sense that it is conceptually and explanatorily independent of other concepts.

The deflationist's "problem of formulation" is now that of finding a theory meeting the above demands.[1] I will assume that there are six candidate types of formulation:

(i) metalinguistic,

(ii) infinitary,

(iii) Ernest Sosa's,

(iv) standard quantified biconditionals featuring propositional quantifiers,

(v) schematic, and

(vi) direct propositional quantificational.

On the first type of formulation, the truth-theory consists in a claim to the effect that the (T)-instances have a certain property, or that some relation holds between their left- and right-hand sides. The second formulation type, argued for by Horwich (1998), shuns any requirement of finite generalisation, and simply takes the correct theory to be the infinitely large "collection" of (non-pathological) (T)-instances. Sosa's formulation (1993) is, "For all propositions *P*, *P* is equivalent to <*P* is true>". The fourth formulation consists in a (first-order) universally quantified biconditional, whose right-hand side contains an existential propositional quantifier, "Σ", as in "$\forall x$(*x* is true iff $\Sigma p$(*x* = <*p*> & *p*))". The fifth type of formulation, proposed by Hartry Field (1994, 2006), is to take (T) itself to be the truth-theory. Finally, the sixth type of formulation goes by prefixing a universal propositional quantifier, "Π", to (T), and treat its schematic sentence-letters as variables, i.e., "($\Pi p$)(<*p*> is true iff *p*)". I will proceed by pointing out flaws or weaknesses of all types of formulation but the last, in

order next to show that the last variant satisfies the directness desiderata perfectly, avoids the problems afflicting the other formulations and is generally attractive. I also argue that the main reason this formulation has been seen as problematic is confused. Thus, the upshot will be that the preferred formulation is better than any other available formulation. But since it is hard to think of further types of formulation, and, more importantly, since the objections against the form of theory I will recommend will be found wanting, there is reason to think that the advertised formulation is also the best formulation *tout court*.

## *II The metalinguistic formulation*

The most obvious idea of how to ensure that the instances of the truth-schema follow from the truth-theory this is to have the theory say simply that these instances are *true* (call this claim (TT)). This would be an instance of the "metalinguistic" type of formulation. However, Horwich (1998: 26f.) has argued that one cannot actually infer the schema-instances from (TT), but only, for each instance, the claim that it is true. In order to infer the instances themselves, we would need certain instances, namely, those where the instantiating sentences are themselves instances of the schema. But, as with all other instances, we cannot infer these instances from the claim that they are true. Now, there is an assumption underlying this argument, which is that an adequate truth-theory must entail the (T)-instances without relying on them. This may seem like a contentious constraint. However, without it, we would get the absurd result that any sentence whatsoever (or at least any tautology) is an adequate theory of

truth. This constraint is thus necessary on pain of trivialising the task before us. It could be motivated by the idea that the truth-theory should "make explicit" what underlies the assertibility of the (T)-instances, or some such, but we need not delve deeper into this, since the constraint is obvious in any case.

Actually, what this reasoning shows is not only that we cannot rely on the (T)-instances when deriving them, but also that their derivation must not *depend*, in a certain sense, on occurrences of "true". Supposing we take as our truth-theory, "(Π*p*)(<*p*> is true iff *p*)", we can see that here, while "true" is used in (the statement of) the theory, and while we can infer the (T)-instances from this theory, this inference does not in the relevant sense depend on the occurrence in the theory of "true". Rather, it is just a case of universal instantiation (with a propositional quantifier). By contrast, an inference of "*p*" from "<*p*> is true" depends essentially on the occurrence of "true". If such derivations were allowed, then, again, we could trivially derive the (T)-instances from any sentence by inferring between sentences of the form "<*p*> is true" and "*p*" and apply conditional proof and equivalence introduction.

Another, related problem with (TT) is that one cannot be taught what "true" means by being given (TT) as an explanation. Such a sentence can equally be taken to implicitly define "false", at least partly, since substituting "true" with "false" would result in an equally true or acceptable claim, namely, that all instances of <<*p*> is false iff *p*> are false. This, one might argue, is not acceptable if (TT) is to be exhaustive. Worse still, since falsity satisfies (TT) as

well as truth does, (TT) does not determine a unique property, and this level of indeterminacy, one might think, is not acceptable (even if *some* indeterminacy perhaps is). Thus, I will assume in what follows that (TT) is not an acceptable truth-theory.

Another proposal of the metalinguistic type might be the claim that we *ought to accept* the (T)-instances. First, we can see that this theory would fare badly with all three directness desiderata. For deriving the (T)-instances from this claim presupposes the instances of "If (one ought to accept <*p*>), then *p*", whence the third directness desideratum would be flouted. There is a further problem with this assumption. Whereas the claim giving wide-scope to "ought", i.e., "One ought to (accept <*p*> only if *p*)", is quite acceptable, the assumption needed, which gives the ought-operator narrow scope, is true only if we read "ought" as expressing the "objective epistemic ought". True, even in the subjective sense, the inference from "One ought to accept <*p*>" to "*p*" is (presumably) correct in the sense that if one ought to accept the former, then one ought to accept the latter. But it is not clearly an *entailment*, since, on that reading of "ought", what one ought to accept may be false.

What about the "objective" reading, on which indeed the claim that one ought to accept <*p*> entails that *p*? Interpreting "ought" in the truth-theory above in this way yields a kind of truth-theory of which there are many other variants, but against which there is a general argument. The type of theory in question is simply one that takes some factive property and ascribes it to the schema-instances. Another example is the claim that the (T)-instances *state facts*. Although such theories entail the (T)-instances, as opposed to the normative theory interpreted subjectively, we clearly could not infer the (T)-instances from

this theory without violating the directness desiderata. In particular, the first one would be violated in that we are relying on non-logical inferences involving the notion of fact, and the third one would be violated in that the theory would entail that the (T)-instances state facts. These points of course hold independently of which factive notion is used. This concludes our case against the first type of metalinguistic formulation, on which we ascribe some property to the (T)-instances.

The other kind of metalinguistic formulation is a claim to the effect that some relation holds between the halves of the (T)-instances. Now, we might consider as a truth-theory the claim that they are derivable from, or equivalent to, each other. This proposal fails for reasons analogous to those adduced against the type of metalinguistic theory above. How, for instance, might "equivalent", as used in this kind of theory, be understood? If it is defined as holding between *x* and *y* just in case: *x* is true just in case *y* is true, then the theory fails for now familiar reasons. Another proposal is that we could define it by the following: <p> is equivalent to <q> $=_{df}$ *p* iff *q*. But this is a schema, and thus does not allow us to infer anything (or, if it does, we might just take (T) as our truth-theory, a proposal we will examine below). And, of course, remedying this by ascribing truth or some other factive property to the schema also fails, for the reasons give above. Quite generally, what we need is a definition of equivalence, given which we can infer "*p* iff *q*" from "<*p*> is equivalent to <*q*>". One proposal is to introduce a function *f* from pairs of propositions to their biconditional, i.e., from <*p*> and <*q*> to <*p* iff *q*> and then propose, "*x* is equivalent to $y =_{df} F(f(x, y))$". From the truth-theory under consideration we may now infer, "$F(f(\langle p\rangle, \langle q\rangle))$" and hence "$F(\langle p$ iff

*q*>)". But for familiar reasons, "*F*" can neither be the truth-predicate, nor any other factive notion. But then, the (T)-instances will not be derivable. If, further, we try a definition of the form, "*x* is equivalent to $y =_{df} F(x)$ iff $G(y)$", the same question arises about "*F*" and "*G*". Somehow, we must infer "*p* iff *q*" from "$F(\langle p \rangle)$ iff $G(\langle q \rangle)$", but this, it seems, can be achieved only with factive predicates (and indeed, using a predicate stronger than the truth-predicate seems to make the derivation of the (T)-instances still more problematic). But it is hard to see what other form a definition with the desired consequence could take. I conclude that this variety of the metalinguistic formulation will not yield an adequate truth-theory.[2]

## *III The infinitary formulation*

Horwich holds that the correct theory of truth is the Minimal Theory (MT), which is defined as the collection of all and only non-pathological instances of (T) (1998: Ch. 1). These "axioms" of (M) are infinite in number (in fact, the cardinality of this infinity is too large for the instances to form a set). Horwich admits that the infinitary character of his theory is a weakness, but he considers it inevitable, given the drawbacks of other formulations (1998: 25ff.).

The most serious problem with Horwich's view is that there is reason to think that we simply *cannot* accept (MT). For to accept a theory is (surely) to accept its axioms, but we simply cannot accept all the axioms of (MT). Even if we could tacitly believe infinitely many propositions, we still cannot believe, even tacitly, all axioms of (MT), since some of them are

too complex to entertain. Take, for instance, the proposition expressed by a (T)-instance whose instantiating sentence contains one thousand negations. Clearly, the vast majority of (MT)-axioms are not entertainable. With the plausible assumption that we should be able to accept the correct theory of truth, we can infer that (MT) is not that theory.

One might now reply that it *is* possible to accept all the instances of (T), for to accept a conditional is merely (roughly) to be disposed to accept the consequent upon (i) acceptance of the antecedent and (ii) consideration of the consequent. However, this identification is false. One can be competent with the conditional connective and have the disposition in question and yet not accept the conditional. This may happen, for instance, if one is irrational or confused enough, or simply on the basis of an "alternative" logical theory (cf. Williamson (2007: Ch. 4)). Similarly, one can accept the conditional without having the disposition. This is precisely what philosophers who reject *modus ponens* (should) do (as Vann McGee (1985) reportedly did). Note that this is weaker than the claim that a competent speaker must, *qua* competent, be disposed to accept a conditional just in case she has the acceptance-disposition is question. I am not denying that, but rather an identity claim as to what it is to accept a conditional. It should make no difference whether we are speaking of accepting sentences or believing propositions. If there are conditional propositions, surely, we can reason fallaciously or "alternatively" with them as we can with sentences (even if possession of the conditional concept requires that we be disposed to reason correctly with it). Thus, the original contention stands: it is impossible to accept (MT), whence it is not the right truth-theory.

But, someone might object, surely we can accept theories like PA or ZFC, and hence,

that a theory has unentertainably complex axioms does not entail that one cannot believe it. I find this completely unconvincing. Surely, the objection confuses different senses of "accepting PA and ZFC". To wit, if these theories are thought of as containing all the instances of their axioms schemata, we can deny the premise of the objection without strain. If we do think there must be a sense in which we can accept these theories (as seems plausible), it is easy to pinpoint such a sense: we can accept these theories, but not conceived of as containing all the instances of the schemata, but rather conceived of as containing some metalinguistic claim (e.g., that all the relevant schema-instances are true) or some claim involving higher-order quantification). If someone still insists that we can accept PA or ZFC in the stronger sense, we can instead object that the instances of (MT) are also *uncountably many*. Thus, unlike the case with PA or ZFC, we cannot decide which its axioms are. This is an additional reason for holding that one cannot accept (MT).[3]

Since many readers have thought the argument above originally occurred in Anil Gupta's (1993a, 1993b) papers on (MT), let me stress that this is not so: Gupta rather stressed the maximal ideology (in Quine's sense) of (MT) (1993b: 365) and some other problems relating to its infinity, but he did not mention the impossibility of accepting it. As far as I know, this argument is new (as are the arguments that follows below).

Let us consider some possible "retreat positions" available to Horwich, i.e., other claims involving (MT) that could be taken as the deflationary truth-theory. Many conceivable theories will just be the kind of theories treated in the foregoing section. For instance, the claim that (MT) is true, though a claim it is possible to accept, fails in requiring the (T)-

instances in order to infer them. Ascribing other factive properties to (MT), or its axioms, as we have seen, is ruled out by the constraints and desiderata. Normative claims fail the same way as above. What about saying that (MT) is the *best* theory of truth? For this to work, we must assume the instances of "If <*p*> is in the best theory of *x*, then *p*". But there is reason to think that the claim that <*p*> is part of some best theory does not properly entail <*p*>. For the instances of "If <*p*> is in the best theory of *x*, then *p*" are not clearly conceptually necessary. After all, they are false on certain, patently coherent, sceptical views. And of course this claim also violates the directness desiderata.

In light of the (perhaps surprising) result that (MT) cannot be accepted, one may wonder, what made it seem that we were dealing with a graspable theory? The answer may be that, in one sense, we do grasp (MT), namely, in the sense that we know the conditions for a proposition's being one of its axioms. The statement of these conditions is itself a finite, general claim. *This* claim, however, clearly cannot be the truth-theory, since it is true by definition and empty.

Once we have duly separated the finite definition of (MT) from (MT) itself, certain further oddities emerge more clearly. In particular, it becomes clearer that the facts expressed by the (T)-instances are primitive and mutually unrelated in an unattractive way. They involve the same property, or concept, of course, but there is nothing in the theory that shows what else they have in common. It can easily *seem* that something shows this, since they are all facts "of the same form", but nothing in the theory mentions this. What ties the facts together can be gathered from the description of the theory, but this description is not part of the theory and does nothing to make the facts in question less primitive and unrelated. As it stands, (MT) is a theory about a certain property, on which one proposition has this property just in case snow is white, and another just in case coal is black, and so on. But, intuitively, a theory of

truth, if there is one, should describe the relevant pattern among the facts, rather than just enumerate them.

This enumerative character of (MT) also casts doubt on its ability to explain facts about truth. First, we must distinguish between a theory, fact, or claim, explaining something, on the one hand, and a *person* explaining something, on the other. Beginning with the latter, it may seem that a person can explain *x* only if she can at least *entertain* that which explains *x* (in the other sense of "explain"). But then, since we cannot entertain the axioms of (MT), we cannot explain the facts about truth. Turning to the other notion of explanation, there is also reason to think that (MT) (or its axioms) cannot explain facts about truth. Firstly, it is unclear whether general facts about truth, e.g., that everything known is true, can be derived from (MT) (see Gupta (1993a: 66) and, e.g., Horwich (1998: 137, 2010: 5.5)). Secondly, it seems to be in the nature of an explanation that *explanantia* essentially involve generalisations or laws, from which the *explananda* can be inferred (sometimes together with further particular facts). If this is correct, then the theory of truth should contain a generalisation, rather than merely particular claims, in order to be explanatory. The type of theory I recommend below has none of these problems, and also allows the most natural way of deriving general facts about truth, which, however, is unavailable to Horwich precisely because (MT) only contains particular claims.

There is one more problem with Horwich's account that the quantificational theory avoids. Consider the theory of facts. Surely, that theory should entail the instances of "That *p* is a fact iff *p*". However, assuming this theory is subject to the same constraints as the truth-theory, it cannot simply be the claim that the instances of these schemata state facts, since this theory would face the same problems as (TT) (that the instances could not be derived unless they are themselves assumed). Could it instead state that the instances of this schema are true? It seems not, for this would then be an alleged fact about truth, and hence something that must

be explained by the truth-theory plus auxiliary claims. But it is hard to see how this claim could be derived from (MT) (or from any other deflationary truth-theory for that matter) plus auxiliary claims. It could of course be derived from the (T)-instances if there were already some other way of deriving the instances of "That *p* is a fact iff *p*" from the fact theory, but the idea was that the latter were to be derived from the claim that they are true. So, the fact theory can say neither that the schema instances state facts, nor that they are true. But then, if plain schemata and propositional quantification are excluded, as Horwich assumes, then it seems the only way the fact theory could be formulated is as an infinite collection of claims. Since there are many other theories that should entail certain schema instances (for instance, our theory of knowledge should entail the instances of, "If *x* knows that *p*, then *p*", and so on), we will have not just one theory with infinitely many axioms, but quite a few of them. If an infinite theory is unattractive in the case of truth (which even Horwich admits), then it must be worse still if the truth-theory commits him to take many theories to be infinite.

## *IV Sosa's theory*

Ernest Sosa (1993) proposes the following claim as a deflationary truth-theory:

(FMT) For all propositions *P*, *P* is necessarily equivalent to <*P* is true>.

(I use capital letters to indicate that the variables are first-order ones, ranging over propositions, rather than propositional variables.) Note that the quantification here is first-order. The obvious problem with this theory concerns the use of "equivalent". For now familiar reasons, it must not be explained in terms of truth or some other factive property. Sosa instead explains this notion in terms of mutual entailment, and explains entailment, in turn, by the schema,

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

Of course, we can infer all the instances of (T) if we have all the instances of (PE). But the problem is that (PE) is a schema. We cannot take the theory of entailment to say that the (PE)-instances are true, or have some other factive property. If can use propositional quantifiers here, further, then we could just use them in the truth-theory as well (which would also yield a more direct generalisation over the (T)-instances). McGrath (1997b) tries to solve this problem by proposing, "For all *P*, *Q*, *P* entails *Q* iff IF(*P*, *Q*) is necessary". This is understood as a first-order quantification over propositions, and "IF" is a function taking pairs of propositions to the conditional proposition from the first proposition in the pair to the second. But how are we to understand "necessary" here? One may suspect that this is merely elliptical for "necessarily true". (Note that the necessity operator is irrelevant here, since what is needed is a predicate applying to propositions, not a sentential operator. However, the suspicion that "necessary", as used by McGrath, involves truth may be understood as the claim that he uses it so that "<*p*> is necessary" expresses the proposition that necessarily(<*p*> is true).) Even if there is a notion of necessity of propositions which does not involve the notion of truth, this solution both conflicts with the directness constraints and falls afoul of the general considerations concerning factive properties of section II.

There is, however, a completely different way of developing Sosa's idea, which is to replace equivalence with *identity*, i.e., to have the theory say,

(F) For all propositions *P*, *P* = the proposition that *P* is true.

This would in fact let us infer the (T)-instances by purely logical means (assuming "=" is a

logical expression). However, (F) is too strong. It entails that to believe that snow is white *is* to believe that the proposition that snow is white is true. Although this case, and certain others, may be considered acceptable, and intuitions to the contrary explained pragmatically (as in Båve (2009)), there is one case that seems impossible to account for in any such way. To wit, there is a widespread intuition that while <snow is white> is true because snow is white, it is not the case that snow is white because <snow is white> is true. (F) contradicts this intuition too, but in this case, there seems to be no promising pragmatic explanation. Also, (F) makes a mystery of propositions involving "blind truth-ascriptions", like the proposition that everything the Pope says is true. So, I conclude that Sosa's type of formulation does not give us an acceptable truth-theory either.

### *V Standard quantified biconditionals featuring propositional quantifiers*

This (fairly popular) type of truth-theory reads,

(TA) $\forall x(x \text{ is true iff } \Sigma p(x = \langle p \rangle \ \& \ p))$.[4]

The quantifier "Σ" (along with its universal counterpart, "Π") are *propositional quantifiers*, whose defining characteristic is that their variables enter *sentence position*, rather than term position. This is not to be conflated with the kind of quantification used in (FMT), where instead a first-order quantifier, ranging over propositions, is used. Clearly, replacing "Σ" with a first-order existential quantifier in (TA) would result in an ill-formed sentence, since the two last occurrences of "*p*" occupy sentence position. Although the use of propositional

quantifiers may seem to be a major worry concerning (TA), I will postpone my discussion about this matter until the section where I present my favoured truth-theory (where I will defend the use of such quantifiers). The real problem with (TA), I will argue, is that it requires a non-logical premise in order for us to infer the (T)-instances. (Inferring the (T)-instances from (TA) also requires a greater number of inference-steps than on my theory, whence the latter fares better with desideratum 2.)

Now, what is the additional premise needed? It is easily seen, given natural inference rules for the propositional quantifier (see, e.g., Prawitz (1965)), that we can derive the instances of "If *p*, then <*p*> is true" from (TA), by universal instantiation, Σ-introduction and some obvious further steps. The converse, however, requires a further assumption or rule. To see this, suppose first we derive, by universal instantiation, "<snow is white> is true iff Σ*p*(<snow is white> = <*p*> & *p*)". But to go from here to the relevant (T)-instance, we would need to infer from "Σ*p*(<snow is white> = <*p*> & *p*)" to "snow is white". In general, we need the instances of the schema, "If Σ*p*(<*q*> = <*p*> & *p*), then *q*" (note that only "*q*" is a schematic sentence-letter here and "*p*" is a bound variable). But these instances are not logical theorems, even if we count propositional quantifiers as logical. Rather, they are true or assertible partly due to the meaning of pointy brackets, which are not logical constants (or at least not quite at the logical end of the continuum, if logicality is a matter of degree—cf. also Künne (2003: 338, 353f.)). (TA) thus violates the first directness desideratum (although it scores rather high overall, compared to the theories previously considered).

Despite these drawbacks of (TA), it may be thought to have an advantage over my preferred formulation in that it gives a condition shared by all and only truths. Firstly, this is not obviously an advantage. A deflationist might for instance argue that the desire to find such a condition stems from the traditional but unwarranted model of concepts as analysable only by explicit definitions. More importantly, by using the schema required to derive the (T)-instances from (TA), we can derive (TA) from the theory I propose, i.e., the claim that ($\Pi p$)(<*p*> is true iff *p*). Thus, that theory would provide the same necessary and sufficient condition for truth. To derive (TA), take an arbitrary term "<*q*>", and suppose that <*q*> is true. The truth-theory entails that if <*q*> is true, then *q*. By *modus ponens*, we infer that *q*. By self-identity and conjunction introduction, we have that <*q*> = <*q*> & *q*. By existential generalisation, we infer that $\Sigma p$(<*q*> = <*p*> & *p*) and, by conditional proof, we have that if <q> is true, then $\Sigma p$(<*q*> = <*p*> & *p*). For the converse inference, assume that $\Sigma p$(<*q*> = <*p*> & *p*). By the appropriate instance of the schema, "If $\Sigma p$(<*q*> = <*p*> & *p*), then *q*" and *modus ponens*, we derive that *q*, which, together with the truth-theory, gives us that <*q*> is true. By conditional proof, the conditional proved above, and equivalence introduction, we get that <q> is true iff $\Sigma p$(<*q*> = <*p*> & *p*). Since "<*q*>" was arbitrary, we infer (TA) by universal generalisation (this last inference step will be discussed at length in the next section).

## *VI The quantificational formulation, Part 1: formal issues*

The quantificational formulation takes the correct truth-theory to be expressed by a sentence formed by prefixing a universal propositional quantifier to (T), i.e.,

(Q) $(\Pi p)(\langle p \rangle$ is true iff $p)$.

Since this is the theory I will be defending here, I should also make clear my view of pointy brackets. To wit, I take the instances of "$\langle p \rangle$" to refer to propositions rather than sentences (see Båve (2009) for a defence and an account of truth-ascriptions to non-propositions). An obvious advantage of this choice is that we immediately avoid Tarski's (1935/1983: 158ff.) and others' well-known objections against propositionally quantifying into quote-contexts. Another important point that should be made here at the outset is that I will not be concerned at all with the attempts by various philosophers of avoiding reference to propositions or avoiding using "true", by recourse to propositional quantifiers (cf. esp. Prior (1971), Ramsey (1927, 1928/1991), and Williams (1976)). I gladly accept both propositions and the use of both propositional quantifiers and "true", and I do not think an adequate truth-theory must provide an eliminative definition.[5]

The main worry one may have about this is that "Π" might not be definable in such a way that (Q) comes out as an acceptable truth-theory, or perhaps not intelligibly definable at all. For instance, "Π" is typically introduced in terms of truth-conditions. On the

substitutional version, for instance, the definition says that a sentence "$(\Pi p)\Phi$" is true iff for all $s$, $(s/p)\Phi$ is true, where "$\Phi(s/p)$" refers to the expression resulting from replacing all free occurrences of "$p$" by $s$ (see Kripke (1976: 330)). In order to infer the (T)-instances from (Q), given such a definition, we clearly need the (T)-instances. Objectual interpretations are of course truth-theoretic too, and thus of no avail, and similarly for semantics using propositional quantifiers in the metalanguage (cf. Hugly and Sayward (1996: Ch. 14) and Williamson (1999)). Another view of quantifiers is that they abbreviate infinite conjunctions or disjunctions (see Field (1984)). Very roughly, we might then say that "$(\Pi p)$ $S(p)$" abbreviates, "$S(p_1)$ & $S(p_2)$ & … & $S(p_n)$", where "$S(\ldots)$" is a sentence-context, and "$p_1$", "$p_2$", …, "$p_n$" are all the sentences of the language. The main problem here concerns "abbreviate". The obvious idea is that it means, "is equivalent with". For now-familiar reasons, there is no way of defining equivalence that will make the resulting account acceptable. If, further, we try to define the quantifiers by inference rules, such as, "$(\Pi q)$ ... $q$ ... $\Rightarrow$ ... $p$ ...", then we are using a schema (cf. Horwich (1998: 26)). Thus, it may seem that there is no way of defining propositional quantifiers given which (Q) comes out as an adequate truth-theory.

I believe this worry is misplaced, for the following two reasons. First, we do not need definitions or inference rules in order for us to understand the propositional quantifiers. We already understand them! Consider the sentences "$(\Sigma p)$ $p$", "$(\Pi p)$ $p$", and "$(\Pi p)$ $x$ knows that $p$ iff $x$ justifiably believes that $p$ & $p$". We can readily see that the first is trivial, the second

absurd, and the third a controversial theory of knowledge. But we could not "see" these things if we didn't understand propositional quantifiers. One clearly does not need any technical definitions or rule-statements in order to understand an expression, even if they may help speakers attain understanding. How, then, did we come to understand propositional quantifiers? Presumably, by analogy with first-order quantifiers (I will sketch such an explanation below). Further support for the intelligibility of propositional quantifiers comes from the examples of non-nominal quantification in English discussed by Prior (1971: 37), Strawson (1974), and Künne (2003: 65ff., 362f.).

Objection: suppose we introduce "Boolean dyadic logical connective" quantifiers "E" and "A" as in, "E*c*(*p c q*)" that stand to logical connectives the way first-order quantifiers stand to terms and propositional quantifiers to sentences. Surely, we cannot understand sentences containing "E" and "A" (except metalinguistically). But since this quantification is relevantly similar to propositional quantifiers, we cannot understand the latter either! Both of the objection's premises can be questioned. First, the premise that we couldn't understand the "connectival" quantification (non-metalinguistically) is not obviously true. One may for instance argue that since connectival quantification can be introduced the same way as first-order quantification (and/or since they have parallel meaning-constitutive principles), the former *must* be intelligible.

Secondly, it is not obviously true that connectival quantification is relevantly similar to propositional quantification. For instance, the "substitution class" of the former is arguably a syntactically arbitrary set, whereas those of first-order and propositional quantification are, respectively, the set of terms and the set of sentences. This might make a difference as to whether we could understand the quantifiers (non-metalinguistically). Even if connectival quantification were introduced so as to acquire a syntactically non-arbitrary substitution class,

say, the set of expressions that can be put between two sentences to get a grammatical sentence, there might still be a difference relevant to our ability to understand. For instance, it may be that we are unable to operate with this syntactic notion the way we operate with the notion of sentence or term at the cognitive level at which the relevant linguistic competence is realised. Perhaps the kind of tacit knowledge needed (for humans) for this competence is such that we cannot know—in the relevant way—propositions involving this notion. I conclude that the objection involving connectival quantifiction is inconclusive.

It may also be objected that although we understand propositional quantifications, we do so in a way that renders (Q) inadequate as a truth-theory. To wit, one may think that "($\Sigma p$) $p$" can only be understood as meaning that *something is true*, thus making (Q) viciously circular. But this involves a contentious assumption regarding logical form. Although the biconditional "($\Sigma p$) $p$ iff something is true" is presumably true, its halves clearly differ with respect to surface grammar (the right-hand side, for instance, containing an ordinary one-place predicate). The objection, however, assumes that they are synonymous, which entails that the surface grammar of the left-hand side is misleading. But imputations of misleading surface grammar always require independent evidence, and I can think of none in the present case. Finally, it is important to note that while the arguments above seem to be wanting, they actually target a claim that is stronger than I need, namely, that we (you and I) in fact understand propositional quantifications as non-truth-involving and non-metalinguistically. But even if (surprisingly, I would say) the reader only understands these expressions this way, surely, they *can* be understood the way I claim to understand them, i.e., in perfect analogy with first-order quantifications. And that weaker claim is sufficient to show that (Q) has a reading (namely, the one I intend here) on which it is adequate as a truth-theory (in the relevant respect).

The second reason why the above worry about (Q) is misplaced is very simple: *we do*

*not need any further premises in order to infer the (T)-instances from (Q), any more than we need further premises in order to infer by universal instantiation in general*. This claim is further sustained by my positive proposal about our understanding of propositional quantifiers, given which the aforementioned analogy with first-order quantifiers will become clear. It says that our competence with "Π" consists, roughly, in

(i) a disposition, for all *s*, to infer from a sentence "(Π*p*)Φ" to "(*s*/*p*)Φ",

(ii) a disposition to accept "(Π*p*)Φ" when, for some *s*, "(*s*/*p*)Φ" can be derived from premisses not containing any expression in *s* (or from no premisses).

I believe that to give the competence conditions of an expression is to give its meaning, but we need not make this specific assumption here. (We could also say, along Horwichian lines, that the fact that we have dispositions (i) and (ii) constitutes the meaning of "Π".) This claim about our competence with "Π" displays obvious parallels with the first-order universal quantifier. The competence condition above, especially (ii), is of course subject to controversy, but this holds equally of first-order quantifiers. Now, competence with the first-order universal quantifier, I propose, consists in

(a) a disposition to infer, for all terms *t*, from "(∀*x*)*Fx*" to "*Ft*"

(b) a disposition to accept "(∀*x*)*Fx*" when, for some term *t*, "*Ft*" can be derived from

premisses not containing any expression in *t* (or from no premisses).

If this is right, it is not surprising how we could have come to understand propositional quantifiers "by analogy with first-order ones", despite lacking a generally agreed-upon, explicit meaning-postulate. I have here committed myself to a broadly "functionalist", "use-theoretic", or "conceptual role" account of semantic competence. But this is no contentious commitment in the context, since deflationism enforces some such account anyway (or so I will assume).

On the present proposal, the inference from (Q) to the (T)-instances is no different from any first-order universal instantiation. It is thus a *logical* inference. If so, then (Q) satisfies the first directness desideratum. The second directness desideratum is also well satisfied, since we can derive a (T)-instance from (Q) by a single inference-step. We also seem to be able to logically derive only the instances themselves (and what follows logically from them), whence the third desideratum is met as well.

As intimated above, it would be confused to require, for inferring in accordance with universal instantiation, that the inference be legitimised by some further principle that must be made explicit. To require that would lead to a regress reminiscent of that famously illustrated by Lewis Carroll (1895). Since we cannot reasonably require this for first-order quantifiers, we should not require it for propositional quantifiers, especially given their parallelism. Thus, the problem for (Q) that Horwich claimed to be insurmountable, namely, that of finding non-

schematic principles required for one to be allowed to infer by universal instantiation, is really a non-problem.[6]

Similarly, to complain that I have not explained fully the meaning of "Π" and that it should not therefore be used in the truth-theory should be no more convincing than the corresponding complaint about "∀". What is important is that we understand "Π", not that we have a correct theory about what its meaning consists in, or what understanding it amounts to. Requiring such an account for all expressions we want to use in stating a theory would force us, absurdly, to scrap virtually every theory we have. Does the fact that "Π" is not part of ordinary English somehow tell against using it? Again, this would similarly tell against many theories we should not want to dismiss, and seems like an absurd demand in any case. It may be thought that first-order quantifiers are "safer" in that they have been around longer and have not given rise to any paradoxes (none serious enough to give us reason to abandon them, anyway). But we cannot be sure that first-order quantifiers are any less paradoxical than propositional ones.

Some think that we have a better-behaved truth-theoretic account of first-order quantifiers than we have of propositional quantifiers, and that using the former is somehow more legitimate than using the latter. Even accepting the controversial premise of this argument, we can respond that since deflationists hold these truth-theoretic accounts to be theoretically otiose, they need not take any such difference to matter for the usability or legitimacy of the respective quantifiers. There are also proof-theoretic accounts of both types

of quantifiers, but they parallel (i)-(ii) and (a)-(b), and thus also indicate that the quantifier-types are on a par (see, e.g., Prawitz (1965)). Quite generally, I take the burden of proof to be on whomever regards propositional quantifiers as different from first-order ones in some way that might be relevant for assessing (Q).

Horwich's second argument against (Q) is that "the use of substitutional [i.e., propositional] quantification does not square with the *raison d'être* of our notion of truth, which is to enable us to do *without* substitutional quantification" (1998: 25). Though this is not completely clear, the idea seems to be that (Q) involves some kind of superfluity. But we have seen that we need propositional quantifiers in any case, since we must not use the truth-predicate in stating various other theories that should entail infinitely many schema-instances. Also, if I am right about what it takes to master propositional quantifiers, they are cognitively cheap, once we have mastered first-order quantifiers.

A different objection might be that if (Q) is the best candidate among the theories considered here, there will be other theories that are equally good. But, one might think, the correct truth-theory should be the uniquely best theory. The best candidates of "equally good" theories might be, firstly, the theory consisting in (T) itself, given an appropriate "definition" of schematic letters (i.e., the schematic formulation). Field has proposed that schematic letters be introduced by way of inference rules (1994: 259) and (2006). (Note that in this paper, schemata have not been used as sentences. Rather, I have referred to schemata only in order to speak of their instances, and in a way that could be eliminated, e.g., in favour of quantification over expressions and concatenation (cf. Båve (2006: 101)).) If (T) is to be a

truth-theory, however, it must be possible for schemata to be used as sentences (as they often are in philosophy and logic). But when they are so used, they arguably function just like universal propositional quantifications. Thus, they will be introduced and governed by the same inference rules. If they are, however, then (T) and (Q) will say the very same thing. Thus, they do not express *distinct*, equally good theories, after all. This argument assumes that on a plausible version of the uniqueness requirement, it should not matter that the sentences are different, if they express the same content. If wording counts, however, then we can argue that (T) is *not* as good as (Q). For while every sentence where "Π" takes widest scope can be replaced by schemata, sentences in which it does not cannot. Thus, since "Π" cannot be replaced by schematic letters without expressive loss, (T) requires the introduction of an otiose notation. Depending on how we individuate theories, then, (T) is either identical or inferior to (Q).

## *VII The quantificational formulation, Part 2: putting (Q) to work*

In this section, I will discuss how (Q) fares with a number of more substantial demands on truth-theories. A major worry facing any truth-theory are the truth-paradoxes. I will argue that the three ways of dealing with these paradoxes are at least *prima facie* open to (Q). Firstly, (Q) seems compatible with the strategy of holding that some alleged sentences or propositions are such that the inference rules allowing us to derive a contradiction (or an arbitrary conclusion, by Curry's paradox) do not apply to them. Thus, one might take some alleged

instances of "<*p*>" to be ill-formed, not referring to a proposition, or some such, and take (Q) to entail only (T)-instances with "legitimate" instances, in some sense to be spelt out. Secondly, the strategy of accepting the unrestricted truth-schema is clearly available, since it just means resting with (Q) as it stands, and adopt a paracomplete or paraconsistent logic. Thirdly, solutions consisting in restrictions of the truth-schema could be adapted to (Q) in the following way. Since all (T)-instances are of the same form, a restriction thereof can be described by reference to the instantiating sentences or propositions. If there is such a restriction, then there must be some predicate *F* which applies to a candidate truth-bearer *x* only if the (T)-instance whose instantiator is *x* is "safe" (where the set of safe (T)-instances is consistent). But then, we could replace (Q) with, "(Π*p*)(*F*(<*p*>) → (<*p*> is true iff *p*))". Thus, all three approaches to the paradoxes seem available to adherents of a (Q)-like truth-theory. While we cannot know that any one of them will succeed, this uncertainty is ubiquitous.

I promised above to show also how a number of problems afflicting Horwich's truth-theory can be solved by (Q). One of these problems is that Horwich must take many theories to be infinite the way (MT) is, for instance, the theory of facts and that of knowledge. On the present account, instead, we can take those theories to be formulated using "Π". Another problem for Horwich was that of deriving general truth-claims. He has shown, for many claims of the form "∀*x*(…*x* is true…)", that one can infer from the (T)-instances every instance of "…<*p*> is true…" (i.e., every sentence resulting form replacing "*p*" here with a declarative sentence). It is more difficult, however, to see how the universal quantification is

to be inferred from these instances.[7]

For (Q), the task is to ensure that first-order quantifications can be inferred from propositional ones. I will try to show here that we can derive general truth-claims from (Q) and auxiliary, truth-free assumptions merely by executing the competence-grounding dispositions described above (in fact, (i) and (b) will suffice). Let us thus see how "(∀*x*) if one knows *x*, then *x* is true" could be so inferred. (This is of course a relatively simple fact, but this should not matter, since it is as good an example as any of the kind of *explanandum* causing trouble for Horwich's account.)

The auxiliary assumption needed here must come from the theory of knowledge. For the reasons given above, this theory must make use of propositional quantifiers instead of "true". If adequate, it will entail, "(Π*p*) If one knows <*p*>, then *p*". Now, from this claim, together with (Q), we can infer "If one knows <snow is white>, then <snow is white> is true", independently of accepting any sentence containing "<snow is white>" (by (i) and equivalence elimination, *modus ponens*, and conditional proof). Now, "<snow is white>" is to be counted as a *term* in the sense of (a) and (b). Thus, on the basis of the derivation of "If one knows <snow is white>, then <snow is white> is true", a competent speaker will accept "(∀*x*) if one knows *x*, then *x* is true", by (b).

Note the difference between (Q) and (MT). The "arbitrariness" required for deriving a universal claim according to the above proposal can never be had on Horwich's theory, since it contains particular instances of (T). Therefore, no inference of a sentence can ever be

derived independently of accepting a sentence containing a particular expression of the form "<*p*>", as required by (b). This point should be clear despite the rough formulations of the competence-grounding dispositions.

All that has been shown here, of course, is that we can infer general truth-claims by executing our competence-grounding dispositions to infer plus auxiliary assumptions. But what was to be shown, it may be thought, was that we would be *right* to derive such claims on the basis of (Q). Well, this is not quite true. If it were, then we could easily achieve the task by adding an assumption to the effect that we are (defeasibly) justified in executing our competence-grounding dispositions to infer.

However, it should be clear by now that it is crucially *not* the case that that assumption is needed to infer "(∀*x*) if one knows *x*, then *x* is true" from (Q) plus the theory of knowledge. Nor, of course, are (i) or (b). To think otherwise is to invite Carroll's regress all over again. To show that "(∀*x*) if one knows *x*, then *x* is true" follows from (Q) and the knowledge-theory, what we must do is *derive it from them*. The first steps of this derivation are obvious by now: we take an arbitrary proposition, say, <snow is white>, and derive from (Q) and our knowledge-theory that if someone knows <snow is white>, then <snow is white> is true. However, there is a problem with the rule corresponding to disposition (b): if we have introduced "proposition" into our language, we will be able to infer that everything is a proposition, since for arbitrary expression of the form "<*p*>", we may categorically assert (or, derive from the null set), "<*p*> is a proposition" (I regard expressions of the form "<*p*>", as

used here, as synonymous with "that"-clauses, so the same holds for the latter). There are in fact a host of other related problems that come with introducing "proposition". This word is introduced so as to validate the schema, "That *p* is a proposition". But thereby, an ordinary rule of existential generalisation will allow is to infer from "That the Sun is shining is nice" to "There is a nice proposition" and from "*X* fears that it will rain" to "*X* fears a proposition". And there are many more examples.[8] If we are to keep "proposition" (in the relevant sense) in our language, then, we must reject the introduction rule for the universal quantifier and replace it. One suggestion is to replace it with the following two:

(UG) Infer "$(\forall x)Fx$" when, for some term *t* not of the form "<*p*>", "*Ft*" can be derived from premisses not containing expressions in *t* (or from no premisses).

(UGP) Infer "$(\forall x)(x$ is a proposition $\rightarrow Fx)$" when, for some term of the form "<*p*>", "*F*<*p*>" can be derived from premisses not containing expressions in "*p*" (or from no premisses).

With (UGP), we could infer the universal claim we wanted (Horwich (1998: 137) proposes a rule that is similarly specific for propositions). Of course, revising our standard rules for the sake of saving a truth-theory would be a bad trade-off. But it is the introduction of

"proposition" that causes the trouble, not the truth-theory. An obvious alternative, then, is to ban "proposition" or perhaps introduce it some way that does not allow us to infer that everything is a proposition with the standard intro-rule for the universal quantifier (note that "true", "false", etc., do not need "proposition" to be inferred, but merely "that"-clauses or pointy brackets). In that case, we keep that rule and use it to infer "($\forall x$) if one knows $x$, then $x$ is true".

Finally, we should also say something about the meaning of "true". I propose we identify giving the meaning of an expression with giving its competence conditions, and, further, that the competence condition for "true" is that one be disposed to infer between the instances of "<*p*> is true" and "*p*". (Thus, I accept an "inconsistency theory of truth", in the sense of Eklund (2002). It is important, however, to note that this theory is terribly ill-named, since it is not a truth-theory, but a theory about our competence with "true".) Note that on this view, (Q) can be derived merely by executing competence-grounding dispositions for "true" and "Π" plus classical logic. To wit, we first infer "snow is white" from "<snow is white> is true" and apply conditional proof. Next, we do the same for the converse conditional, apply equivalence introduction, and finally Π-introduction. While problems with paradoxes may seem to surface at this point, I will (perhaps opportunistically) leave them for another occasion.

**Acknowledgement** I would like to thank the participants at the Language and logic seminar at Stockholm University, especially Peter Pagin and Dag Prawitz, as well as to two anonymous referees at *Synthèse*.

## REFERENCES

Armour-Garb, B. (2010), "Horwichian Minimalism and the Generalization Problem", *Analysis* **70**, pp. 693-703.

Asher, N. (1987), "A Typology of Attitude Verbs and Their Anaphoric Properties", *Linguistics and Philosophy* **10**, pp. 127-189.

Bach, K. (1997), "Do Belief Reports Report Beliefs?", *Pacific Philosophical Quarterly* **78**, pp. 215-241.

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

Båve, A. (2006), *Deflationism: A Use-Theoretic Analysis of the Truth-Predicate*, Acta Universitatis Stockholmiensis, doctoral dissertation, Stockholm, Almqvist & Wiksell.

Båve, A. (2009), "Deflationism and the Primary Truth-bearer", *Synthèse* **173**, pp. 281-297.

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

Carroll, L. (1895), "What the Tortoise Said to Achilles", *Mind* **4**, pp. 278-280.

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

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

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

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

Eklund, M. (2002), "Inconsistent Languages", *Philosophy and Phenomenological Research* **64**, pp. 251-275.

Evans, G. and J. McDowell (eds.) (1976), *Truth and Meaning*, Oxford, Clarendon Press.

Field, H. (1984), "Is mathematical knowledge just logical knowledge?", *The Philosophical Review* **93**, pp. 509-552.

——— (1994), "Deflationist Views about Meaning and Content", *Mind* **103**, pp. 249-285.

——— (2006), "Compositional Principles versus Schematic Reasoning", *The Monist* **89**, pp. 9-27.

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

Gupta, A. (1993a), "Minimalism", *Philosophical Perspectives* **7***: Language and Logic*, pp. 359-369.

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

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

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

Heidelberger, H. (1968), "The Indispensability of Truth", *American Philosophical Quarterly* **5**, pp. 212-217.

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

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

——— (2010), *Truth-Meaning-Reality*, Oxford, Oxford University Press.

Hugly, P. and C. Sayward (1996), *Intensionality and Truth*, Dordrecht, Reidel.

Kalderon, M. E. (1997), "The Transparency of Truth", *Mind* **106**, pp. 475-498.

King, J. (2002), "Designating Propositions", *Philosophical Review* **111**, pp. 341-371.

Kirkham, R. (1992), *Theories of Truth*, Cambridge, MA., MIT Press.

Kneale, W. (1972), "Propositions and Truth in Natural Languages", *Mind* **81**, pp. 225-243.

Kovach, A. (1997), "Deflationism and the Derivation Game", *Mind* **106**, pp. 575-579.

Kripke, S. (1976), "Is There a Problem about Substitutional Quantification?" in Evans and McDowell (1976), pp. 325-419.

Künne, W. (2003), *Conceptions of Truth*, Oxford, Oxford University 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.

McGee, V. (1985), "A Counterexample to Modus Ponens", *Journal of Philosophy* **82**, pp. 462-471.

McGee, V. (2000), "The Analysis of '*x* is true' as 'For any *p*, if *x*="*p*", then *p*'", in Chapuis and Gupta (2000).

McGrath, M. (1997a), "Weak Deflationism", *Mind* **106**, pp. 69-98.

——— (1997b), "Reply to Kovach", *Mind* **106**, pp. 581-586.

Moltmann, F. (2003), "Propositional Attitudes without Propositions", *Synthese* **135**, pp. 77-118.

Pap, A. (1952), "Note on the 'Semantic' and 'Absolute' Concept of Truth", *Philosophical Studies* **3**, pp.1-8.

Platts, M. (1979), *Ways of Meaning*, London, Routledge and Kegan Paul.

Prawitz, D. (1965), *Natural Deduction: A Proof-Theoretical Study*, Stockholm, Almqvist & Wiksell.

Prior, A. N. (1971), *Objects of Thought*, Oxford, Clarendon Press.

Raatikainen, P. (2005), "Horwich's Way Out", *Analysis* **65**, pp. 175-177.

Ramsey, F. (1927), "Facts and Propositions", *Proceedings of the Aristotelian Society, supp. vol.* **7**, pp.153-170.

Ramsey, F. (1928/1991), *On Truth*, ed. Rescher, N. and Majer, U., Dordrecht, Kluwer.

Resnik, M. D. (1990), "Immanent Truth", *Mind* **99**, pp. 405-424.

Schantz, R. (ed.) (2002), *What is Truth?*, Berlin, de Gruyter.

Schiffer, S. (2006), "Propositional Content", in E. Lepore and B. Smith (eds.), *Oxford Handbook of Philosophy of Language*, Oxford, Oxford University press, pp. 267-294.

Soames, S. (1999), *Understanding Truth*, Oxford, Oxford University Press.

——— (2003), "Understanding Deflationism", *Philosophical Perspectives, 17, Language and Philosophical Linguistics*, pp. 369-383.

Sosa, E. (1993), "The Truth of Modest Realism", *Philosophical Issues* **3**, *Science and Knowledge*, pp. 77-95.

Strawson, P. F. (1974), "Positions for Quantifiers", in M. K. Munitz and P. Unger (eds.), *Semantics and Philosophy*, New York, New York University Press, pp. 63-79.

Tarski, A. (1935/1983), "The Concept of Truth in Formalized Languages", translated and reprinted in A. Tarski, *Logic, Semantics, Metamathematics*, Indianapolis, Hackett.

Tarski, A. (1944), "The Semantic Conception of Truth", *Philosophy and Phenomenological Research* **4**, pp. 341-376.

van Inwagen, P. (2002), "Generalizations of Homophonic Truth-sentences", in Schantz (2002), pp. 205-222.

Williams, C. J. F. (1976), *What is Truth?*, Cambridge, Cambridge University press.

Williamson, T. (1999), "Truthmakers and the Converse Barcan Formula", *Dialectica* **53**, pp. 253-270.

Williamson, T. (2007), *The Philosophy of Philosophy*, Oxford, Blackwell.

[1] Discussions about this issue can be found in Tarski (1935/1983, 1944), Pap (1952: 8, n. 2), Heidelberger (1968), 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 (1994: 267f.) and (2006), Kalderon (1997), Kovach (1997), McGrath (1997a, b), Horwich (1998: 25ff.), Soames (1999: 42, 86ff.), McGee (2000), Dodd (2000: 36ff.), van Inwagen (2002), Hill (2002: 17ff.), and Künne (2003: Chs. 2 & 4). In my own doctoral dissertation *Deflationism* (2006), I argued for a number of claims concerning this problem, many of which I have here abandoned.

[2] Soames (1999: 230) and (2003: 372) proposes as a deflationary theory the claim that the propositions expressed by instances of "The proposition that *p* is true" and "*p*" are *a priori* consequences of each other, which he takes to mean, roughly, that it is possible in principle to reason deductively from one to the other without appeal to empirical evidence. But of course, it is not the mere *possibility of deriving*, but the possibility of deriving *validly*, that matters. If this qualification is included in the theory, however, then, since validity must be spelt out in terms of truth (or possibly some other factive notion), the problems discussed above will arise.

[3] This objection against my argument and the alternative argument involving uncountability are due to an anonymous referee at *Synthèse*.

[4] (TA) is discussed in Ramsey (1928/1991: 9), Kneale (1972), Mackie (1973), Baldwin (1989), David (1994: 74ff.), Kalderon (1997: 491), van Inwagen (2002), Hill (2002: 22), and Künne (2003: 6.2). McGee (2000) discusses analyzing "*x* is true" as "For all *p*, if *x* = '*p*', then *p*".

[5] It may be thought that my attitude violates some sound principle of parsimony in that propositional quantifiers on the one hand, and using "true" or admitting propositions, on the

other hand, are exchangeable (according to the philosophers engaged in the relevant eliminative projects). But I will argue below that propositional quantification is cognitively very cheap, and it arguably comes at most with an ontological commitment to propositions, to which I am already committed. As for the eliminability of "true", propositional quantifiers and ordinary English (minus "true") may well be insufficient for paraphrasing all sentences with "true" (consider, "Most things he says were true but aren't anymore"). If not all English sentences with "true" can be paraphrased using only propositional quantifiers and English, this is presumably because "true", being an ordinary adjective, can be combined with many other expressive devices ("most", tensed copulas, etc.) in ways propositional quantifiers cannot (cf. Båve (2006: 4.3)).

[6] In Båve (2006: 3.3), I persisted in similar fallacies, as I was (vainly) forewarned by Peter Pagin and Dag Prawitz.

[7] His most recent attempt to deal with this problem is in his (2010: Ch. 5, sect. 5), criticised by Armour-Garb (2010). See also Raatikainen (2005) and Field (2006).

[8] These problems relating to "proposition" have been discussed by Asher (1987), Bach (1997), King (2002), Moltmann (2003), Künne (2003), Schiffer (2006), and Båve (2006: 5.6).