

# WHAT IS TRUTH?

*Arhat Singh Virdi*

**Submitted For Degree of Ph.D. in Philosophy**

**London School of Economics and Political Science  
Houghton Street, London  
WC 2A 2AE**

**2010**

I declare that the work presented in this thesis is my own.

The copyright of this thesis rests with the author. Quotations from it are permitted, provided that full acknowledgement is made. This thesis may not be reproduced without prior written consent of the author.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Word Count (including footnotes but excluding bibliography and appendices): 57, 919

# ABSTRACT

I defend the correspondence theory of truth, according to which a statement's truth consists in a relation of correspondence with extralinguistic fact. There are well-known objections to this view, which I consider and rebut, and also important rival accounts, principal among which are so-called *deflationist* theories and *epistemic* theories. Epistemic theories relate the concept of truth to our state of knowledge, but fail, I argue, to respect the crucial distinction between a criterion of truth and the meaning of truth: the view that one cannot do semantics, or metaphysics, without addressing epistemic issues is rejected by this work. Against epistemic theories, I illustrate how truth is independent of epistemic considerations. Deflationism is the more popular of the rival accounts and has gained considerable momentum over the past two decades. It is therefore dealt with in greater detail by this work. Deflationist theories exploit the paradigmatic “‘Snow is white’ is true iff snow is white’ biconditional to argue for an insubstantialist account, according to which truth is conservative with respect to non-semantic facts. On this view, truth's *raison d'être* is merely to perform the useful expressive function of generalising over possibly infinite sets of assertions. Against deflationist theories, I claim that the work done by Jeffrey Ketland and Stewart Shapiro conclusively demonstrates how truth is informationally additive over non-semantic facts, while deflationism itself is also an excessively impoverishing theory, inadequate to the tasks it purports to accomplish.

This work also defends the thesis that Alfred Tarski's well-known theory of truth is an authentic correspondence theory. To say this is to say that the clauses of a Tarskian truth-definition can be interpreted in terms of a relation of correspondence that holds between true sentences and the states of affairs they describe. I provide a precise account of what the correspondence in question consists in, claiming that true sentences are homomorphic images of facts, i.e. a true sentence represents, in a form-preserving manner, the truth-making facts in it. This gives precise expression to Wittgenstein's thesis that true sentences picture the world.

# CONTENTS

ABSTRACT.....	3
CONTENTS.....	4
PREFACE.....	7
CHAPTER 1. INTRODUCTION.....	8
CHAPTER 2. EPISTEMICISM.....	16
2.1 Motivating Epistemicism about Truth	
2.2 Coherentism	
2.3 Peircean Pragmatism	
2.4 Jamesean Instrumentalism	
2.5 The Non-Viability of Epistemic Theories of Truth	
2.6 Appendix	
CHAPTER 3. CORRESPONDENTISM.....	36
3.1 The Primal Attraction of the Correspondence Theory	
3.2 The Correspondence Relation	
3.2.1 Deflationism and the Correspondence Relation	
3.3 What are Facts?	
3.3.1 The Slingshot Argument	
3.3.2 Is Correspondentism the View of Metaphysical Realism?	
3.3.3 Taking Stock	
3.4 Ramsey's Ladder	
3.5 Dr. Tarski, Or: How Truth Theorists Learned to Stop Worrying and Love Semantics	
3.5.1 Correspondentist Aims	
3.5.2 Truth is Relational	
3.5.3 Satisfaction	

3.5.4 Truth-in-a-Model	
3.5.5 Tarskian Truth and Facts	
3.5.6 Addressing Three Popular Objections	
3.6 Closing Remarks	
3.7 Appendix	
CHAPTER 4. DEFLATIONISM.....	79
4.1 Introductory Remarks	
4.2 Deflationary Theory Not Committed to the <i>T</i> -scheme	
4.2.1 Strawson's Illocutionary Theory	
4.2.2 C.J.F. Williams' Redundancy Theory	
4.2.3 Prosententialism	
4.3 <i>T</i> -scheme Based Deflationism	
4.3.1 Redundantism	
4.3.2 The Central Objection to Redundantism	
4.3.3 Disquotationalism	
4.3.4 Minimalism	
4.3.5 Objections	
CHAPTER 5. DEFLATIONISM AND CONSERVATIVENESS.....	120
5.1 The Conservativeness Implication	
5.1.1 An Analogy with Hilbert's Programme	
5.2 Is Deflationary Truth Conservative?	
5.2.1 Tennant's Rebuttal	
5.3 Appendix	
CHAPTER 6. CONCLUSION.....	136
BIBLIOGRAPHY.....	143



# PREFACE

I would like to thank Colin Howson for his careful supervision of my thesis, for reading (uncountably) many drafts of each chapter and for guiding my style. His advice has been invaluable. I would also like to thank Hannes Leitgeb for his charitable help and patience in allowing me to see ways in which the thesis presented can be improved. Naturally, I would like to thank my brother, Adipat, and my sister, Arhan, for their unfailing love, support and inspiration.

Gurcharan Kaur, my mother, passed away 6<sup>th</sup> November 2002 just after my third year as a PhD student at the LSE had begun. My mother suffered severe hardships throughout her life, yet her spirit remained unbroken as she remaining firmly wedded to the virtues of forgiveness, love and tenderness toward others. Whosoever came into contact with her could not walk away without being greatly impressed by my mother's gentility and magnanimity. Lamenting her lack of education<sup>1</sup> – citing her not possessing a university degree and her sufferance as evidence – my mother persevered to ensure her children's educational well-being. All caring parents wish the best for their children. In my mother's case, this was everything. I have often thought about this. My mother taught English to non-native speakers and organized and administered courses for a highly successful London college for adult learners (in her honour that college annually awards a prize to the student who overcomes the greatest difficulties in winning an ESOL certificate). My mother singly raised, nurtured and schooled three children while maintaining a home, its finance and under painfully difficult circumstances. These are feats (my experience of) people schooled even in the very top universities would be unequal to. They serve as a testimony to my mother's very unique fortitude and conviction. They also served to strengthen my own resolve in continuing with this effort when all was lost that autumnal day. Contrary to what she herself thought, my mother's life is a beacon of light. My mother is the only angel I shall ever know; she is my *sine qua non*. I dedicate this effort to my mother.

---

<sup>1</sup> My mother studied for an MSc in Education (at the Institute of Education, University of London) but the need for her to be hospitalized from the cancer very sadly prevented her from completing the degree.

# CHAPTER 1

## INTRODUCTION

Logic, like any science, has as its business, the pursuit of truth. What are true are certain statements; and the pursuit of truth is the endeavour to sort out the true statements from the others, which are false.

**W.V.O. Quine, *Methods of Logic***

When entering upon the study of a science, we need to have some idea, if only a provisional one, of its nature. We want to have in sight a goal to strive towards; we want some point to aim at that will guide our steps in the right direction. The word 'true' can be used to indicate such a goal for logic, just as can 'good' for ethics and 'beautiful' for aesthetics. Of course all the sciences have truth as their goal, but logic is concerned with the predicate 'true' in quite a special way, namely in a way analogous to that in which physics has to do with the predicates 'heavy' and 'warm' or chemistry with the predicates 'acid' and 'alkaline'. There is, however, the difference that these sciences have to take into account other properties besides these we have mentioned, and that there is no one property by which their nature is so completely characterized as logic is by the word 'true'.

**Gottlob Frege, *Logic***

This work seeks to defend the twin claims that truth is, in nature, a substantial notion and, in meaning, correctly articulated by the correspondence theory of truth. To understand what these claims are, and why they are considered correct, one must undergo a sort of propaedeutic of what a philosophical investigation of truth consists in. Let us begin by making clear that the thesis defended here does not hinge on how we are to understand the medium through which truths get conveyed: our truth-vehicles could be beliefs, ideas, judgements, propositions, statements, sentences, other forms of linguistic representation. For the sake of argument, let us here take propositions to be our truth bearers. Consider, now, the question 'What is truth?' What exactly is this thing, truth, that propositions are endowed with an ability to carry? Common amongst the philosophical community is an expression of bafflement concerning truth. This might be a consequence of the Fregean view expressed above that there is no one property characterizing any science so completely as 'true' does logic. Thus, no other pieces of information seem available, as one might expect in the other sciences, to help illuminate the concept.<sup>2</sup> Quine, however, did not share this view. He thought the concept of truth is *unambiguous*:

There are philosophers who stoutly maintain that 'true' said of logical and mathematical laws and 'true' said of weather predictions and suspects' confessions are two usages of an ambiguous term 'true'...<sup>3</sup> What mainly baffles me is the stoutness of their maintenance. What can they possibly count as evidence? Why not view 'true' as unambiguous but very general, and recognize the difference merely between logical laws and confessions?

---

<sup>2</sup> This is part of Frege's reason for holding truth to be *indefinable*.

<sup>3</sup> Quine cites Gilbert Ryle (Ryle 1949a, p.29) and Bertrand Russell (Russell 1912a, Chapter IX) as examples of such philosophers.



(Quine 1960, p. 131)

The reason Quine gave for holding this view is important. They are to do with our pre-theoretical, pre-analytic grasp of truth. But before we proceed to investigate what this might amount to, it is important to get right what the question is asking, and resist the invitation from some corners to apologize for asking it. Some take ‘What is truth?’ to be the *wrong question* to ask about truth.<sup>4</sup> For example, in *An Introduction to Philosophical Logic*, A.C. Grayling argues that this is owing to the question having “the character of a smooth sheer cliff, which one cannot see how to climb..., [it] looks like a request to know what truth (‘Truth’) is in some ultimate, inclusive, perhaps mystical sense.”<sup>5</sup> Granted – but appearances can be deceiving. And as no one would deny Grayling’s sensible counsel “that there can be no progress in understanding a concept unless one’s enquiry begins with the right question”,<sup>6</sup> what, then, would be the *right question* to ask about truth? Grayling’s manner of asking the right question is: “What is it for a proposition to be true?” On this there is no disagreement between us. Our enquiry ‘What is truth?’ is taken to mean precisely this. Truth is a concept that applies to propositions. Clearly, truth is not a concept applying to *all* propositions; it applies only to those that are true. Thus, this investigation seeks to understand ‘What is it that all true propositions share that all false propositions lack?’, or, equivalently, ‘What is it for a proposition to be true?’ i.e. ‘What is truth?’

Still, a lack of clarity might persist. A way of getting to grips with the nature of this enquiry is to understand what it is not. This enquiry is *not* ‘What is true?’; it does not aim to discern truths.<sup>7</sup> It attempts rather to discern what it is about true propositions that constitute their truth.<sup>8</sup> In his seminal *The Problems of Philosophy*, Bertrand Russell gave the following warning (reminiscent of Grayling’s counsel above) to the wayward truth investigator conflating the two questions:

...for the present we ask only ‘What is truth?’ ...not ‘What [propositions] are true?’ ...It is very important to keep these different questions entirely separate, since any confusion between them is sure to produce an answer which is not really applicable to either.

---

<sup>4</sup> Note that this is quite different from the deflationary attitude to the question (see below). According to the deflationist, this is an acceptable question to ask, but any answer involving ‘substantial’ notions would be the wrong answer.

<sup>5</sup> Grayling 1997, p. 122.

<sup>6</sup> *Ibid.*

<sup>7</sup> As it is often misconstrued whenever I respond, ‘Truth’, to the question, ‘What are you writing your Ph.D. thesis on?’

<sup>8</sup> Cf. Davidson: “It is an error to think that if someone seeks to understand the concept of truth, that person is necessarily trying to discover important general truths about justice or the foundations of physics. The mistake percolates down to the idea that a theory of truth must somehow tell us what, in general, is true, or at least how to discover truths.” (Davidson 1999, p. 105).

(Russell 1912a, p. 69)

The question that is the focus of our investigation parallels the type of question focusing Socrates' attention in the Platonic dialogue *Euthyphro* – that of unpacking the notion of 'piety':

*Socrates:* And what is piety, and what is impiety?

*Euthyphro:* Piety is doing as I am doing; that is to say, prosecuting anyone who is guilty of murder, sacrilege, or of any similar crime - whether he be your father or mother, or whoever he may be - that makes no difference; and not to prosecute them is impiety...

...

*Socrates:* ...at present I would rather hear from you a more precise answer, which you have not as yet given, my friend, to the question, 'What is "piety"?' When asked, you only replied, 'Doing as you do, charging your father with murder'.

*Euthyphro:* And what I said was true, Socrates.

*Socrates:* No doubt, Euthyphro; but you would admit that there are many other pious acts?

*Euthyphro:* There are.

*Socrates:* Remember that I did not ask you to give me two or three examples of piety, but to explain the general form which makes all pious things to be pious. Do you not recollect saying that one and the same form made the impious impious, and the pious pious?

*Euthyphro:* I remember.

*Socrates:* Tell me, what is the nature of this form, and then I shall have a standard to which I may look, and by which I may measure action, whether yours or those of anyone else, and then I shall be able to say that such and such an action is pious, such another impious.

...

*Euthyphro:* Piety, then, is that which is dear to the gods, and impiety is that which is not dear to them.

*Socrates:* Very good, Euthyphro; you have now given me the sort of answer which I wanted...

(Hare & Russell 1970, pp. 41-42)

{Needless to say, Socrates goes on to quiz the content of Euthyphro's proposal of what constitutes piety!}

This investigation is not easy. As Plato makes clear in the *Theaetetus*, truths do not come with a 'mark', like a date on the corner of photographs that distinguish them from falsehoods. To many, however, this attitude is puzzling. For them there is a straightforward answer to the question 'What is truth?' Appealing to Aristotle, they are apt to remark that:

To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.

(Aristotle, *Metaphysics*, Γ, 7, 1011<sup>b</sup>25)

True propositions, then, tell it like it is, or (in other words) for a proposition to be true is for it to correspond with the facts. It is in this that the unambiguity of truth resides and to which Quine alludes in the quotation above. It is important to understand that this is a *platitude* that no one denies. As Ralph Walker puts it:

Colloquially, “corresponds with the facts” can function as a long-winded way of saying “is true”... (Walker 1999, p. 309)

*but*, continues Ralph Walker, any theory of truth “so understood...becomes an empty tautology”. Owing to their being harmless, these platitudes do not seem to raise the theoretical temperature very high.<sup>9</sup> So any theory of truth worth its salt must proffer a genuine *elucidation* of the notions involved here.

Many argue that we need not look very far. Latent in platitudes like ‘truth is correspondence with fact’ or ‘truth is agreement with reality’ is a substantial theory of truth, the *Correspondence Theory of Truth*. According to it, the platitude indicates truth’s involvement in a *real* two-place relation (a seriously dyadic one),<sup>10</sup> between propositions, on the one hand, and facts, on the other. Unpacking the notion of truth in this way, it argues, offers a genuine account of truth, one where the notion plays a significant, *informationally additive*, role, i.e. to say of a proposition that it is true is to say something more than that already contained in the proposition itself. This thesis seeks to defend the correspondentist’s point of view.

What could possibly be wrong with the correspondence theory? One difficulty, often cited, lies in the very attempt to flesh out the notions involved in the correspondence aphorism.<sup>11</sup> What *kind* of thing is a fact? Are there negative facts, conditional facts, disjunctive facts, or are facts just ‘general’? And if so, what are they made of? Can they be enumerated? It is not entirely clear that the relevant facts can be specified beyond making the uninformative (and circular) claim that they are those which make a particular proposition true. Moreover, would we not be making the same unwarranted move Russell is often accused of making in reifying universals by saying that true propositions are those which fit the facts? The term ‘fact’ in sentences like ‘The fact that grass is green’ appears to acquire its meaning by reference to some worldly correlate of the proposition ‘grass is green’, but only by presupposing some version of the unstable ‘Fido’-Fido principle: is not the impression that facts are independent existents a consequence of our illegitimately projecting

---

<sup>9</sup> This way of putting the point comes from Simon Blackburn. See, for example, Blackburn 1984, p. 225, or Blackburn 2005, Chapter 3.

<sup>10</sup> This is Crispin Wright’s way of putting the point. See Wright 1992.

<sup>11</sup> William James, for example, is a partisan of this concern observing that “...*quarrels [begin] only after the question is raised as to what may precisely be meant by the term ‘agreement’, and what by the term ‘reality’, when reality is taken as something for our ideas to agree with.*” (James 1907b (1999), p. 53).

onto the world divisions we find operating in thought?<sup>12</sup> Also, what *kind* of correspondence is involved? Arguably, the only kind of understanding involved in saying that true propositions correspond with the world, that they somehow ‘picture’ it, is the metaphorical kind. Surely there is more involved here in cashing the idea of correspondence out than that involved in the employment of a metaphor? But again, it is not entirely clear that anything can be said about the correspondence relation, other than making the uninformative, circular claim that it is the relation with which a proposition stands to the world when it is true. For some, attempts at explaining the notions of fact and correspondence generate nothing but pseudoexplanations – they are understood only in terms of what they are supposed to explain.

Such concerns have made many sceptical about the correspondentist’s ability to offer credible answers to these questions. Pumping the platitude up into a substantial theory of truth involving a two-place relation generates (what are considered to be) insurmountable problems. The theory also seems to require that we tolerate a (not altogether outrageous) separation between the mind and the world. Being naturally (though, as Chapter 3 will highlight, erroneously) associated with metaphysical realism – the view that there is an objective reality whose nature is independent of our language and thought – weakens the theory by lending it to the criticism of epistemological scepticism. Being epistemically unconstrained, the concept of truth is considered to be objective owing to its independence from justification and warrant; it hinges only on the way the world is. Truth is “radically non-epistemic” (to use Hilary Putnam’s words) or “evidence-transcendent” (to use Michael Dummett’s term for realism). The charge, then, is that we can never determine whether propositions are true because we cannot compare them with the facts true propositions are said to correspond to. Such “realist theories” of truth, as Davidson prefers to call them:

...seem to throw in doubt not only our knowledge of what is “evidence-transcendent”, but all the rest of what we think we know, for such theories deny that what is true is conceptually connected in any way to what we believe.  
(Davidson 1990, p. 299)

We cannot confront propositions with facts that are external to them as there is no way of getting outside our language so as to facilitate the comparative examination correspondentism requires. Strongly motivated by this epistemological concern, Carl Hempel spoke of the “fatal confrontation

---

<sup>12</sup> Since the name ‘Fido’ gets its meaning by referring to an individual – Fido – should we suppose *all* words function in a similar way? The ‘Fido’-Fido principle was coined by Gilbert Ryle in his review of Carnap 1947 (Ryle 1949b). Ryle’s attitude to such speculation is actually shared by Oscar Wilde. In *The Importance of Being Earnest*, Gwendolyn Fairfax is made to exclaim: “Ah! That is clearly a metaphysical speculation, and like most metaphysical speculations has very little reference to all the actual facts of real life, as we know them.”

of statements and facts”.<sup>13</sup> The reason typically advanced for why a proposition cannot be compared with the facts to see if a fact corresponding to the statement obtains is that propositions are comparable *only* with other propositions. Initially voiced by Otto Neurath, Carl Hempel and the early Carnap, the objection continues that any attempt at such a comparison with the world results solely in the acquisition of *more* propositions:

We call a content statement “false” if we cannot establish conformity between it and the whole structure of science...The verification of certain content statements consists in examining whether they conform to certain protocol statements; therefore we reject the expression that a statement is compared with ‘reality’, and the more so, since for us ‘reality’ is replaced by several totalities of statements that are consistent with themselves but not with each other.<sup>14</sup>  
(Neurath 1934 (1983), p. 102)

Abjuring the correspondence theory, but still motivated by a concern to subserve the ‘substantiality’ of truth, philosophers attempt to articulate alternate proposals promising to be more faithful to our epistemic situation. The result of such attempts is called *Epistemicism* about truth. Chapter 2 aims at undermining epistemic theories of truth.

Indeed, given the potential difficulties with the correspondentist view indicated above, one might be tempted to think that there is, in fact, something wrong in the *very nature* of the project in discovering what truth is. Many have been tempted by this thought. They invite philosophers to reject abstract questions about the nature of truth, expressing their view by saying that truth is not a ‘real’ or substantial property, or a robust and metaphysically interesting notion. *Deflationism* about truth is the doctrine of those desiring to deflate the inflationary intentions and metaphysical excesses of correspondentism (and epistemicism); they categorically reject the presupposition that truth has an analyzable and reducible ‘inner nature’. Why? We have already encountered a deep concern with the (perceived) generation of pseudoexplanations in attempting to elucidate the

---

<sup>13</sup> Hempel 1935, p. 51. Hempel is not articulating the *representative* view here of the logical positivists’ stance on truth. Though Otto Neurath and Rudolf Carnap (at one time) were co-conspirators, Moritz Schlick, by contrast, was an ardent defender of the correspondence interpretation in his debates with Neurath and Hempel (see Schlick 1935). Karl Popper, too, found no difficulty with the correspondence theory of truth, recording in his intellectual autobiography his belief that Alfred Tarski had rehabilitated “the much maligned correspondence theory” (Popper 1974b, p. 78). And, of course, the manifesto of logical atomism – *Tractatus Logico-Philosophicus* – was written by a philosopher wholly imbued with correspondence intuitions.

<sup>14</sup> Desirous of having a test for truth, Carnap, Schlick et al (in virtue of their strong empiricist inclinations) called statements reporting immediate perceptual experience “protocol statements”, viewing them as incorrigible in that they are directly subject to verification. The truth of other statements is to be tested by their logical relations to these ‘protocols’. Neurath, in fact, was dubious about the supposed incorrigibility of protocols (in a way that is hidden from this quote), and in denying the possibility of their being subject to direct verification, held that the *only* test for truth consists in relations among statements themselves. Scheffler 1967 has a lively account of the controversies between Hempel, Schlick, Carnap and Neurath.

relation of correspondence and the notion of fact. Deflationists surmise that this renders both empty. They argue that the moral to draw from this is that the correspondence account should therefore be trimmed (or ‘deflated’ of naïve correspondence intuitions), leaving bare the insubstantiality of truth. Consider, for example, the proposition that snow is white. According to the correspondentist’s account:

- (1) The proposition that snow is white is true just in case the proposition that snow is white corresponds to a fact

However, presumably:

- (2) The proposition that snow is white corresponds to a fact just in case snow is white

In recognizing this connection, philosophers such as Frege, Ramsey and Quine, proceed to shave (1) as follows:

- (1\*) The proposition that snow is white is true just in case snow is white

(1\*) is a perfectly acceptable fact about what the truth of the proposition that snow is white consists in. It is owing to the triviality of such statements – schematically represented by the Tarskian *T*-scheme – that deflationists urge their tautologousness. In the case above, deflationists continue, the truth predicate is simply superfluous. However, the deflationist *also* recognizes that truth is not universally redundant. It proves to be useful in performing certain linguistic roles. For example, how else could one express one’s agreement with, or faith in, everything that follows from the Peano-Dedekind axioms for arithmetic without employing truth? Of course, one could try to list all the propositions but this would be an infinite task, taking an infinite amount of time to convey the intended meaning. By saying, instead, ‘Everything that follows from the Peano-Dedekind axioms for arithmetic is true’, truth endows one with an ability to *finitarily* express and capture the intended meaning. Truth is here increasing our expressive power.

There are differing views on what deflationism actually is, and is best characterized as a philosophical *tendency* than a specific doctrine. This does not make it less of a challenge however. Arguably, it presents the most potent challenge to correspondentism. While the motivation of the epistemicist’s activity is the epistemological concern to render truth humanly accessible, which they deny correspondentism is able to do, both deflationists and correspondentists nevertheless agree on

where to begin. Both attempt to do justice to our pre-analytical grasp of truth inherent in such platitudes as ‘truth is correspondence to fact’. Correspondentists attempt to fine-grain our already coarse-grained understanding and bring to bear a substantial theory of truth; deflationists, on the other hand, staunchly deny the validity of so doing.

According to the deflationary point of view, once we have identified the theoretically ‘cool’ platitudes associated with it and the logico-linguistic functions it performs, there is no ‘property’ of truth existing independently of these facts. In particular, we need not burden truth with any metaphysical baggage or intumescence. If truth is dispensable and contentless, then one might reasonably suggest that adding a correct theory of truth, call it *T*, to some non-semantical e.g. physical theory of the world, call it *W*, would produce a *conservative extension* – i.e., any (non-semantical) fact provable from the union  $T \cup W$  should already be provable without the addition of *T*. If, as the deflationist contends, truth is insubstantial then, as Stewart Shapiro and Jeffrey Ketland point out,<sup>15</sup> it is difficult to see how she could not to be committed to truth’s conservativeness. It is owing to a commitment of this sort that the deflationist’s position proves to be untenable. There are, for example, deflationary functions that the truth predicate is said to perform which *cannot* be undertaken by a conservatively extending theory of truth. More significantly, conservatively extending theories of truth are inadequate to the task of legitimizing truth claims.

Chapter 2 addresses the epistemological motivation for rejecting the correspondence theory of truth and finds it indefensible. Chapter 3 recounts Alfred Tarski’s *semantic* conception of truth and argues in favour of the view that it exonerates correspondentism. Chapter 4 critically surveys the various brands of deflationism and identifies their various shortcomings. Chapter 5 deals with the issue of deflationism and the conservativeness implication, concluding that there are important features to truth which deflationism is inadequate to the task of justifying, unlike Tarski’s semantic defence of correspondentism.

---

<sup>15</sup> Shapiro 1998; Ketland 1999.

# CHAPTER 2

## EPISTEMICISM

What is the difference which actually constitutes the truth or falsehood of a belief? I am not asking for what is called the criterion of truth, i.e. for some quality, other than truth, which belongs to whatever is true and to nothing else. This distinction between the nature of truth and a criterion of truth is important, and has not always been sufficiently emphasized by philosophers.

**Bertrand Russell, *Philosophical Essays***

According to all forms of epistemicism about truth, truth is an epistemically constrained concept not capable of being evidence-transcendent. Each holds that the proposition that *S* is true means that the proposition participates in some epistemic property *P*, like *S* coheres with some specified set of propositions believed, or *S* is useful to believe, or *S* is that opinion to which investigators will converge in the limit of inquiry. This chapter outlines three popular versions – coherentism (the most popular), pragmatism and instrumentalism – and finds them all indefensible as theories of truth (they may well be viable as theories of rationality). I deal in detail with early instances of epistemicist truth (those of Neurath, Peirce and James) and find that the fundamental criticism that they all fail to obey Russell's (above) instruction of keeping the nature of truth and the criterion of truth distinct is taken to imply that truth is not synonymous with any epistemic property whatsoever. As the features of the early instances I criticize are the same in both the early and late instances of epistemicist truth, there would be no real gain in dealing separately with the later instances – such as Dummett's theory of warranted assertibility, or Putnam's theory of (idealized) rational acceptability.

### 2.1 Motivating Epistemicism about Truth

When people first start to sympathize with the correspondence theory, they are apt to seize upon some homely immediate truth, a truth which "leaps to the eye". I consider, say, the fact that my typewriter is on the table and say 'well, *that's* a fact, *that* is what my judgement that the typewriter is on the table is made true by'. This seems as pure a case of sheer acquaintance with a fact as can be got...The instinct behind such a choice of case is sound. The critic is pressing for a conception of 'the facts' [and of 'correspondence'] which is not wholly derivative from an antecedent conception of true judgement. (Blackburn 1984, p. 233. *Parenthetical addition is mine.*)

Invoking the notion of 'correspondence' and of 'fact' imports no explanatory value to correspondence-type theorising about truth because these *explananda* cannot be understood without appealing to the *explanandum*. The critic is troubled by the fact that nothing more can be said about the correspondence relation *other than* its being the relation that a proposition stands to the world



when it is true. Similarly, it appears that nothing more can be said in unpacking the notion of fact *other than* it is that which makes a proposition true. The generation of these interdependent, circular definitions significantly undermines the sustainability of understanding the notion of truth in correspondence terms. According to Richard Schantz:

It is the concept of truth that wears the breeches; facts and correspondence being simply projecting from truth, according to the very simple rule that the fact that *p* is that which corresponds to the true statement that *p*... Thus all we seem to gain by introducing facts and the relation of correspondence is the ability to say, in esoteric language, that a statement is true.  
(Schantz 2002, p. 2)

Tied to the critic's concern of the impossibility of adequately reducing truth along correspondence lines is the epistemological concern of the impossibility of performing the kind of language-world comparison that the theory is inviting us to make.<sup>16</sup> In order to determine whether a proposition is true we would have to ascertain whether it corresponds to the relevant fact in the appropriate way. But this is impossible, so the thinking goes, as there is no pure, unmediated apprehension of facts as they are in themselves *apart from the way we judge them to be or from our ways of conceptualizing them*. Unlike the cartographer who can check his map against the landscape, we cannot check our linguistic map against the extralinguistic landscape.<sup>17</sup> We cannot escape the 'circle of our beliefs' and scrutinize reality itself; in fact, we are perennially pinned inside it as all our cognition of the world is mediated by our thought and experience. The correspondence theory is implicitly committing us to (the myth of) 'the given'<sup>18</sup> – the uncontaminated, raw presentation of fact cleanly received by our minds – proposing an intolerable gulf between mind and the world, one that, for our critic, simply does not exist.

Those seduced by this epistemological concern need to be careful, however, not to overstate this point. The fact that we can never know whether a judgement corresponds to the relevant facts in the appropriate way does not mean that truth understood as correspondence is wrong. If judgements cannot be known to correspond to reality then we cannot infer that they do not so correspond. Someone committed to the correspondence point of view can still stand their ground. They might even be compelled to agree that extraction of judgement from apprehension of fact is impossible and *still* this would take nothing from the view that truth consists in correspondence between

---

<sup>16</sup> See Ewing 1934 for an early modern introduction to this epistemicist idea.

<sup>17</sup> To make this point, Wittgenstein was fond of a German saying which translates as "you cannot shit higher than your arse" (reported in Blackburn 2005, p. 57).

<sup>18</sup> The phrase "Myth of the Given" is taken from Wilfrid Sellars 1997 *Empiricism and the Philosophy of Mind*.

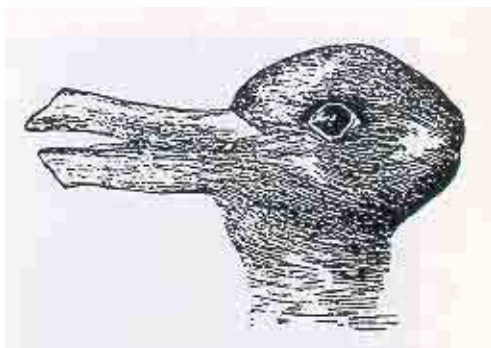
judgement and the relevant facts. But the critic presses for a more radical position: the very notion of fact is itself *dependent* on our judgement:

Even such a low-grade judgement as that my typewriter is on the table involved recognizing that the elements of the situation are spatially external to me, that they are objects with a temporal history, that they have various physical properties such as solidity, and so on. Judgement just *is* the isolation of facts. (Blackburn 1984, p. 234)

As they are carved out by us in language, there can be no objective facts to make acquaintance of. The correspondentist is left impotent, unable *even* to support her claim that we are able to survey a correspondence between perceptual awarenesses of our immediate environment, like seeing a typewriter in front of me – a “pure a case of sheer acquaintance with a fact as can be got”. The idea that our source of conceiving facts cannot be wholly extracted from the operation of judgement-making is typical of Kant, a philosopher many epistemicists cite as their inspiration. Central to Kant’s thought is a critique of the notion of *adaequatio intellectus et rei* – the adequacy of thought or the intellect to things. He sought to expose the interdependence of the categories of thought and experience by saying:

Without sensibility no object would be given to us, without understanding no object would be thought. Thoughts without content are empty, intuitions without concepts blind. It is, therefore, just as necessary to make our concepts sensible, that is, to add the object to them in intuition, as to make our intuitions intelligible, that is, to bring them under concepts. These two powers or capacities cannot exchange their functions. The understanding can intuit nothing, the senses can think nothing. (Kant 1781 (1993), p. 75)

Arguably, the best illustration making the interdependence of experience and thought most apparent are double-aspect figures. Consider, for example, the figures below:



From one point of view the image that leaps out is that of a duck; from another, that of a rabbit.<sup>19</sup> This ambiguous, bistable figure is not illusory; it is not designed to trick the eye. Depending on your direction of attention, there are two kinds of perceptual phenomena legitimately extractable from this one visual stimulus. Changing from one interpretation to the other gives the spectator entirely different visual experiences. Thus, when confronted with facts, we cannot divorce an understanding of their constitution from the products of our own imagination, interpretation and excogitation.

It is owing to this that leads many epistemicists to stress that in seeking to apprehend these extracognitive/extralinguistic facts with which to check correspondence, we end up, in spite of ourselves, with further beliefs or judgements instead; our beliefs are simply not controlled or constrained by the way things are.<sup>20</sup> A problem with the argument from scepticism to a theory in which truth consists in relations between judgements alone is that this *fails* to solve the problem of scepticism. The sceptical thesis (one implicit in Berkeley's famous claim that an idea can be like nothing but an idea) is that what we take to be facts bear no resemblance to mind-independent facts: since we cannot occupy a view from nowhere to observe a correspondence obtaining between our language and (extralinguistic) reality, access to this reality is mediated by our cognition precluding access to this reality as it is in itself. Hence, we can never be justified that our beliefs are accurate reflections of a mind-independent world. The epistemicist's response to the sceptic's concern is to redefine truth so as to make it something attainable. They aim to argue for our having adequate justification for our beliefs about a world which consists of mental constructs not resembling the mind-independent world at all: the phenomenal facts supposedly bearing no cognizable relation to mind-independent facts are artefacts of the epistemicist's imagination. But this fails to deal with the sceptic's worry that our beliefs may not be justified as accurate reflections of an external world. The epistemicist needs to put herself to the task of proving that there *really* is no such world.<sup>21</sup> Davidson made this point too:

Epistemic theories are sceptical in the way idealism and phenomenalism are sceptical; they are sceptical not because they make reality unknowable, but because they reduce reality to so much less than we believe there is.

(Davidson 1990, pp. 298-299)

---

<sup>19</sup> The American Psychologist, Joseph Jastrow, and *not* Ludwig Wittgenstein (as people commonly believe), was the first to note the significance of this double-aspect figure (see Jastrow 1899).

<sup>20</sup> See Blanshard 1941, pp. 227-229 for an argument to this effect.

<sup>21</sup> Cf. Kirkham 1992, pp. 112-114. Kirkham sees clearly that a theory of truth is *not* the place to deal with the problem of scepticism – that would be a burden for theories of justification.

More importantly for our purposes here, a theorist about truth should not be concerned to discover that all epistemic access to facts is indeed mediated by judgement. A theory of truth is not a theory of how facts are accessed epistemically. The latter resides within the province of epistemology while the former in semantics and metaphysics, and the claim that one cannot conduct semantics or metaphysics without also being forced to attend to epistemic issues is to be rejected.

## 2.2 Coherentism

Owing to the belief in the incomparability of judgements with anything apart from other judgements (a belief already undermined), epistemicists seek to outline a theory of truth which renders its understanding wholly reliant on the relationship between judgements alone; nothing other than judgements and their inter-relations would be adequate in attempting to elucidate the nature of truth. The chief motivation has been anti-sceptical, to make truth accessible. Given that no arbitrary set of claims and relations among them will do, philosophers following this line of thought naturally resort to expressing this reliance in terms of how well the judgements ‘hang together’, ‘mutually support each other’ or ‘cohere’. This view found sympathy with elements of the Vienna Circle. As the correspondence theory of truth was seen to postulate independently subsisting facts or states of affairs, articulating suspicious metaphysical premises, it comes as no surprise that that theory got a poor press among the logical positivists. A typical expression of this suspicion comes from Otto Neurath:

Statements are compared with statements, not with ‘experiences’, not with a ‘world’ nor with anything else...<sup>22</sup> Each new statement is confronted with the totality of existing statements that have already been harmonized with each other. A statement is called correct if it can be incorporated in this totality. What cannot be incorporated is rejected as incorrect. Instead of rejecting the new statement, one can alter the whole existing system of statements until the new statement can be incorporated... There can be no other ‘concept of truth’ for science.  
(Neurath 1931 (1983), p. 53)

What does it mean for a body of statements to be coherent? In “Protocol Statements”, Neurath offers the idea of mere consistency: given a set of statements that are true, any addition can be incorporated and thus determined true just in case the set “remains consistent if the statement is

---

<sup>22</sup> This is highly questionable. In logic, the valuation function is a *homomorphism* between syntactical items and their interpretation. For example, consider the term  $t(x_1, \dots, x_n)$ . Then, the value,  $val[t(x_1, \dots, x_n)] = val(t)[val(x_1), \dots, val(x_n)]$ . Of course, this is merely an oversimplified model, but it shows that it is perfectly possible, using the device of a metalanguage (not appreciated by Neurath), to define a relation of representation between language and a structure external to it. I shall come back to this in the following chapter.

added”.<sup>23</sup> If coherence equals consistency then we have an unsustainable theory of truth for the very simple logical reason that any consistent set of statements can contain falsehoods (or, better, that there are consistent sets mutually inconsistent with each other): another way of putting this is to point out the fact familiar to every logician that any consistent set *S* can be extended to two consistent sets *S'* and *S''* obtained by adding an independent sentence and its negation respectively to *S*. So the maintenance of consistency cannot be a guarantor of truth as any new addition could maintain the consistency of the set to which it is added and yet still be false. Understood as such, this would constitute a comprehensive death-blow to the coherentist conception of truth. Moreover, we should not forget the following sober Russellian comments:

[w]hen I say ‘the sun is shining’, I do not mean that this is one of a number of sentences among which there is no contradiction; I mean something which is not verbal, and for the sake of which such words as ‘sun’ and ‘shining’ were invented. The purpose of words, though philosophers seem to forget this simple fact, is to deal with matters other than words. If I go into a restaurant and order my dinner, I do not want my words to fit into a system with other words, but to bring about the presence of food. I could have managed without words, by taking what I want, but this would have been less convenient.  
(Russell 1940, pp. 140-141)

But perhaps we are being too quick in passing judgement on the coherence theory. It is not the case that a partisan of coherentism “must consider any arbitrary fairy tale to be no less true than a historical report or the propositions in a chemistry-book, so long as the tale is well enough fashioned to harbour no contradiction anywhere”.<sup>24</sup> In all its forms, the coherence in question is not just the mutual support that a body of statements give each other, but the mutual support between statements that are *actually believed* or *subscribed to*. The coherence in question is then with a certain designated set of beliefs. A fabricated tale might be a consistent body of statements but that does not mean it is a consistent body of truth-claims. The intended set of statements is the set of beliefs actually held, those which the believer has found reasons for committing herself to, and it is their consistency that the coherentist is offering as the determinant of truth.<sup>25</sup> But then this raises the question of what the grounds are for holding these beliefs. Would it not be that they are true?<sup>26</sup> If so, coherentism is presupposing the concept of truth in its attempt at illuminating it.

---

<sup>23</sup> Neurath 1932 (1983), p. 95. Neurath fails to appreciate that there are uncountably many, mutually inconsistent, extensions of any given set of sentences. So coherence cannot just mean consistency.

<sup>24</sup> Schlick 1934 (1979), p. 376.

<sup>25</sup> Cf. Walker 1999, p. 310.

<sup>26</sup> Possibly not. A Bayesian would say *mutually reinforcing probabilities* (see Bovens & Hartmann 2003), in which case truth is not being presupposed.

In any case, even with the above qualification in mind, it appears that the consistent-yet-false objection remains untouched. The logical objection to viewing truth as coherence, understood to mean consistency, holds. Suppose that the potential entrant to the set of beliefs  $\{p, p \rightarrow \neg q\}$  is  $q$ . By *modus ponens*, the augmented set would be inconsistent: there is no assignment of truth-values to the particles that make up the sentences of this set that makes them all true. Thus, we could either maintain a commitment to the old set, thereby rejecting  $q$ , or revise the initial set of beliefs to enable our being able to accommodate the newcomer in a consistent way. However, in doing the latter, at least one of the sets  $\{p, q\}$  and  $\{p \rightarrow \neg q, q\}$  would have to contain a falsehood. And so Schlick's objection continues to hold. Mutual consistency is simply too weak: there are uncountably many ways of extending a consistent set which are mutually inconsistent with each other.

The Viennese contingent were not the only ones engaged in attempting to elucidate truth as coherence. In fact, in wishing to defend coherentism about truth, one might well do better than appeal to the logical positivists' descriptions for want of not weakening their case by the argument given above. Alluding to the British Idealists of the late nineteenth and early twentieth centuries, it is clear that, perhaps because of appreciating the kind of objection given by Schlick above, coherence was not understood in terms of consistency alone. For example, in "The Nature of Truth" Harold Joachim acknowledges on behalf of the coherentist that:

[t]he 'systematic coherence' ...in which we are looking for the nature of truth, must not be confused with the 'consistency' of formal logic. A piece of thinking might be free from self-contradiction, might be 'consistent' ...and yet it might fail to exhibit that systematic coherence which is truth.  
(Joachim 1906, p. 170)

One suspects that Joachim does not wish 'formal logic' to be interpreted in too limited a fashion here. Not only would a set of formally inconsistent sentences be ruled out as participating in the systematic coherence involved, but also those containing couples that are conceptually inconsistent or together impossible. I have in mind pairs such as 'Colin is a bachelor' and 'Colin is a married man', or 'Peter is the father of James' and 'James is the father of Peter'.

Coherence, then, requires something more than consistency, and this was recognised by the British Idealists. F.H. Bradley, for example, required that a set of sentences needed to be 'comprehensive' *in addition* to being consistent for it to be coherent:

In speaking of system I mean always the union of these two aspects [i.e. consistency and comprehensiveness], and this is the sense and the only sense in which I am defending coherence...neither of these aspects of system will work by itself.

Comprehensiveness deals with the explanatory scope of a set of beliefs: a set of beliefs is more comprehensive than another just in case it has all the explanatory scope of the latter but harbours at least one more description or explanation of a phenomenon that the latter does not. Now, it is important to ensure that the set of beliefs *are* coherent, that they mutually support each other. If we consider the set {Arhat has two eyes, LSE is a part of the University of London, All bachelors are unmarried men}, it contains nothing but truth claims *and* it is more comprehensive than any of its subsets, but it can hardly be considered a set of claims that in any plausible sense cohere. What is required is a sense of coherence in which the members of the set mutually support each other very much like how the poles in a tepee support each other (*cf.* Küne 2003, pp. 384). Blanshard claimed that the poles in this tepee are co-entailments: that each judgement is entailed by the others (either jointly or singly).<sup>27</sup> But a moment's reflection would tell us that this construal is absurd. The best chemistry text book or depiction of a historical period would fail to meet this condition; and so they would fail to be true. But we would hardly want to conclude this, unless of course there were reasons to do so independently of its lack of coherence. What we need is a weaker sort of mutual support, a weaker kind of inferential link. It seems reasonable to suggest that these poles might better be understood as some sort of probability conference: that any subset of the set in question confers a higher probability to a claim than the claim would do of its own accord. Following Küne,<sup>28</sup> the following set of judgements would be good example of this kind of unification between the elements: {Most of the people here are happy, and Ben is here if and only if Ann is here, Ben is here and is happy, Ann is here and is happy}. Assuming any two members of this set confirms the third in the sense that the probability of the third being true is thereby increased.

This would be a good place to gather our thoughts concerning how best to articulate a coherence theory of truth: a judgement is true just in case it is a member of a consistent, comprehensive and (not necessarily logically) mutually entailing set of beliefs. Let us call such a set a *maximally coherent set of beliefs*.

Could not a maximally coherent set of beliefs contain falsehoods? Equivalently, could it not be possible for a proposition to be a member of a maximally coherent set of beliefs *independently* of its truth? Could it not be possible to be epistemically deluded about some object in a fashion that is so

---

<sup>27</sup> Blanshard 1941, pp. 264-265.

<sup>28</sup> Küne 2003, p. 384.

systematic as to escape non-conformity with the properties of a maximally coherent set? Maximally coherent sets can be arrived at without any regard to the actual world; our flights of fancy can generate such sets. Borrowing from Chisholm,<sup>29</sup> Künne says that, after all, “the most hopeless form of insanity is that in which the various factors of the delusion are most systematically rationalised with reference to one another.”<sup>30</sup> Consequently, could it not be possible to acquire more than one maximally coherent set of beliefs concerning an area of discourse that totally diverge from one another? It is their attempt at dealing with this that leaves one considering coherentist views quite frustrated, for the following reason. Coherentists try to dispel concerns relating to systematically delusional states<sup>31</sup> offering sets of maximally coherent beliefs by arguing that what keeps a check or control on these sets is their footing in reality – which, in order to be a check must be independent of what is believed about it. This sounds very much like correspondence talk! In particular, coherentists tend to refer to perceptual judgements acquired from one’s immediate environment – such as ‘I see a typewriter on the table in front of me’ or ‘It is raining’. These are meant to act as a constraint on the maximally coherent set of acquired beliefs. Any maximally coherent set of beliefs must allow for these. Those that do not cannot be true. Bradley, for example, wrote:

With regard to the world as perceived...my power is very limited. I cannot add to this world at discretion and at my pleasure create new and opposite material. Hence, to speak broadly, the material here is given and compulsory, and the production of what is contrary is out of my power.  
(Bradley 1909 (1994), p. 215)

...imagine my world made on the principle of...accepting mere fancy as fact. Could such a world be more comprehensive and coherent than the world as now arranged?...[you must] include everything to be gained from immediate experience and perception...  
(Bradley 1909 (1994), p. 213)

Neurath made a similar point:

...for us striving after knowledge of reality is reduced to striving to establish agreement between the statements of science and as many protocol statements as possible.  
(Neurath 1934 (1983), p. 107)

For Neurath, protocol statements report deliverances of immediate sensory experience. They express, in effect, Bradley’s ‘world as perceived’. The coherentist is going back on herself here. The primary (epistemological) motivation for rejecting truth as correspondence was that in

---

<sup>29</sup> Chisholm 1977.

<sup>30</sup> *Ibid.*

<sup>31</sup> Or indeed any other state that churns out maximally coherent sets of beliefs which have no regard for the way the world actually is.



attempting to judge the nature of any correspondence between a judgement and its associated state of affairs the judger is found, according to the coherentist, to be perpetually trapped by the generation of further judgements and beliefs disallowing her to perform the relevant assaying task. By appealing to protocol statements or to perceptual judgements (that report sensory experiences derived from one's immediate environment) to guard against the objection of systematic delusions the coherentist severely weakens her case. Conditions stipulating when immediate sensory reports are to be included and when not do not seem to be guided by principle. When such reports are meant to play a significant role and when not is ad-hoc. The coherence theory of truth is incoherent.

Further criticism comes from acknowledging that it is clearly possible to have more than one set of judgements that cohere in the relevant fashion. Where to thence? If there is more than one set of beliefs that determine the standard of coherence, then there is no guidance on which one set is to be preferred, which one in whose coherence truth consists. Also, false beliefs can participate in coherent systems; one can imagine a world no less coherent than our own in which a false belief passes as true. But the most damaging criticism of the coherentist position is that it simply fails to supply an adequate theory of truth. The coherence theory may be an adequate theory of rationality, but in offering a theory of truth it fails to appreciate the crucial distinction between the criteria with which to determine whether a proposition is true and the meaning of the concept 'true', and to respect further that a theory of truth is seeking the latter. This difference is very much like appreciating the difference between fixing the meaning of 'feverish' as having a temperature higher than some given point and specifying the procedures for deciding whether someone *is* feverish.<sup>32</sup>

### **2.3 Peircean Pragmaticism**

In wishing to 'make our ideas clear' and banish philosophical arguments whose resolution could have no practical significance to our lives, Peirce was committed to the following pragmatic maxim:

Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object.  
(Peirce 1878 (2001), p. 202)

In application to the concept of truth, Peirce understood the pragmatic maxim to imply that truth is a property of those beliefs unassailable by doubt:

---

<sup>32</sup> Cf. Haack 1978, p. 88.

...if by truth and falsity you mean something not definable in terms of doubt and belief in any way, then you are talking of entities of whose existence you can know nothing, and which Ockham's razor would clean shave off. Your problems would be greatly simplified, if, instead of saying that you want to know the "Truth", you were to say that you want to attain a state of belief unassailable by doubt.  
(Peirce 1905 (1960), p. 279)

The act of having obtained these indubitable beliefs registers that we have come to some form of settled opinion or consensus and reached the end of inquiry. These beliefs on which everyone agrees is the ineluctable conclusion of those who persist enough with their investigations, with all the relevant information, and is independent of the methods they might happen to employ:

If a general belief...can in any way be produced, *though it be by the faggot and the rack*, to talk of error in such belief is utterly absurd.  
(Peirce 1960, Volume 8, Section 16. Emphasis added.)

This is crucial to Peirce's view: if some form of mass hypnosis were just as effective at realising a consensus as that produced by employing the scientific method then mass hypnosis is no less deficient at realising truth. The aetiology of the consensus is irrelevant. By *definition*, a proposition is true just in case it is universally assented to.<sup>33</sup> We are fated to reach truth in the limit of our investigations:

The opinion which is fated<sup>34</sup> to be ultimately agreed to by all who investigate is what we mean by truth...  
(Peirce 1960, Volume 5, Section 407)

Peirce did, however, *also* express the view that, despite our being fated to ultimate agreement once the relevant experiences have been felt and suitably thorough investigations undergone, there is no guarantee that a consensus will be reached.<sup>35</sup> Hence, the final opinion representing the real is one that only *might* be reached *if* all the relevant investigations are carried out to their limit. {This observation will prove fatal for Peirce later}. Nonetheless, when Peirce does express confidence in the fact that all those who engage in the relevant investigation *will* eventually reach a limit in which

---

<sup>33</sup> Although, Peirce did emphatically believe that the scientific method is the most efficient in facilitating universal assent.

<sup>34</sup> Peirce adds the following footnote here: "*Fate means merely that which is sure to come true, and can no how be avoided. It is a superstition to suppose that a certain sort of events are ever fated, and it is another to suppose that the word fate can never be freed from its superstitious taint. We are all fated to die.*"

<sup>35</sup> Cf. "*I do not say that it is infallibly true that there is any belief to which a person would come to if he were to carry his inquiries far enough. I only say that that alone is what I call Truth. I cannot infallibly know that there is any truth.*" (Peirce 1966, p. 398).

everyone agrees, it is owing to a doctrine he held concerning reality. According to Peirce, reality is objective and mind-independent. Owing to this, its impact on our individual perceptions is such that reality determines what is perceived; we cannot control or influence it. This fact is true for everyone investigating. Hence, everyone will reach the same conclusion at the end of their investigations:

The conception of truth gradually develops...reaching the idea of truth as overwhelmingly forced upon the mind in experience as the effect of an independent reality.  
(Peirce 1960, Volume 5, Section 564)

So far, there seems very little to distinguish the theory of truth unfolding here and the correspondence theory: doesn't the plausibility of this pragmaticist view lie in the plausibility of the correspondence theory? There appears to be no further light to be shed with the Peircean identification of 'is true' with 'is ultimately agreed to by everyone investigating' than that already shed by the correspondentist's identification of 'is true' with 'corresponds to the facts'. But Peirce held a curious doctrine concerning reality. Reality is objective and mind-independent of any finite number of minds, not mind-independent of *all minds*. This is what makes the pragmaticist's view on truth epistemic: there is no allowance made for a reality which is in principle epistemically unattainable and plays no role in guiding our inquiry. The opinion ultimately agreed on by all those who investigate is what represents the real – the above quotation proceeds thus: "...and the object represented in this opinion is the real. That is the way I would explain reality". Elsewhere, Peirce outlines his 'social theory of reality' thus:

Reality is independent of , not necessarily of thought in general, but only of what you or I or any finite number of men may think about it...On the other hand, though the object of the final opinion depends on what that opinion is...what that opinion is does not depend on what you or I or any man thinks.  
(Peirce 1960, Volume 5, Section 408)

This view is untenable. Peirce is arguing that a reality independent of any one mind (or of any finite subset of minds) impinges on us externally causing us to undergo certain experiences. As it is mind-independent, this reality will force all those involved in an investigation to undergo similar experiences in a fashion that, in the limit, will produce a consensus opinion. But that reality is an idealist object: it is that idea in all those minds that have followed through with their investigations to the bitter end. How can a group of people undergo experiences whose causal origin is in a reality determined by the end product of their exercising their cognitive faculties over those very experiences? Peirce does admit that it seems paradoxical to claim that "the object of final belief

which exists only in consequence of the belief, should itself produce the belief”.<sup>36</sup> But in defending himself against this fundamental weakness, Peirce claims that there is nothing extraordinary:

...in saying that the existence of external realities depends upon the fact, that opinion will finally settle in the belief in them. And yet that these realities existed before the belief took rise, and were even the cause of that belief, just as the force of gravity is the cause of the falling of the inkstand – although the force of gravity consists merely in the fact that the inkstand and other objects will fall.  
(Peirce 1960, Volume 7, Section 344)

The analogy Peirce is offering in his defence is inadequate. First, he is stating the relation between gravity and the tendency for objects to fall to be one of identity alone. Second, even if we accept this, gravity and objects falling are not temporally locatable events in the way that the occurrence of perceptions that mould individual beliefs and the opinion arrived at by all in the limit of investigation are. Thus, asserting that gravity is both the cause and consequence of objects tending to fall does not validate the possibility of reverse-chronological causation.<sup>37</sup> The analogy Peirce is suggesting breaks down.

A further observation undermining this view of causation is that if, as Peirce admits, the opinion settled on in the final outcome is one of hypothetical existence – it is not “infallibly true that there is any belief to which a person would come to if he were to carry his inquiries far enough” – then how can that opinion representing the real possibly cause our present actual experiences? The thesis advanced here is that some *possible* existent in the future *actually* shapes our current experiences. One ought to seriously doubt entertaining the plausibility of a theory of truth so based.

## 2.4 Jamesian Instrumentalism<sup>38</sup>

In James’s writings, one can find endorsements of a verifiability theory, a coherence theory, a Peircean consensus theory and a correspondence theory. His response to critics pointing this out is that they “seem...to labour under an inability almost pathetic, to understand the thesis which they seek to refute”,<sup>39</sup> claiming quite outrageously that “[t]he critics have boggled at every word they

---

<sup>36</sup> Peirce 1960, Volume 7, Section 340.

<sup>37</sup> I borrow this way of identifying this position from Kirkham 1992.

<sup>38</sup> John Dewey and F.C.S. Schiller held an instrumentalist view on truth closely allied to James’s. For example, see Dewey 1911 and Schiller 1907.

<sup>39</sup> James 1909, p. 10.

could boggle at, and refused to take the spirit rather than the letter of [the pragmatist's] discourse".<sup>40</sup> Nevertheless, it is the instrumentalist view which James is most popular for espousing and is the focus of this section.

James was united with Peirce under the pragmatic maxim – "[t]here can *be* no difference anywhere that does not *make* a difference elsewhere"<sup>41</sup> – but was led to abandon the consensus view Peirce saw as a logical consequence of the maxim. James simply denied the existence of some ultimate opinion obtainable in the limit of inquiry, claiming that:

...the fact that all experience is a process, no point of view can ever be the last one. Every one is insufficient and off its balance, and responsible to later points of view than itself.  
(James 1909, p. 55)

In sympathy with this Jamesean point of view, Richard Rorty poses the following question to the Peircean: "How would we know that we were at the end of inquiry, as opposed to merely having gotten tired or unimaginative?"<sup>42</sup> By responding that the end of inquiry is signalled by some convergence in opinion *does not* posit a time-invariant belief. James's counsel not to expect some endpoint to inquiry is a result of his conception of judgement as tools of our own making, tools designed to equip us with the means to deal with our surroundings. He gets to this conception by the following route. First, James endorses the pre-theoretical, correspondence-type view on truth: as he himself said, that conception is one you would find expressed in any good lexicon (*cf.* James 1907b (1999), p. 53). But James is swayed by the sceptical concern of how we could ever be justified in the truth of our beliefs if reality was mind-independent:

How does the partisan of absolute reality know what this orders him to think? He cannot get direct sight of the absolute; and he has no means of guessing what it wants of him except by following the humanistic clues.  
(James 1909, p. 46)

James appropriates the Kantian outlook that there is no route out of our own conceptual scheme to see if the sensations we receive are products of a mind-independent world at work. Our perceptual organs and concept-deployment organize our experiences. Our ideas about objects and the objects themselves are indistinguishable (*cf.* James 1909, p. 35). Reality is, therefore, "an accumulation of

---

<sup>40</sup> *Op. Cit.*, p. 99.

<sup>41</sup> James 1907, p. 30.

<sup>42</sup> Rorty 1986 (1999), p. 329.

our own intellectual inventions”.<sup>43</sup> But, unlike Kant, James does not think that such constructs are built into our minds. They are, rather, inventions made by our ancestors, inventions designed in the way they are because it proved useful to do so:

[A]ll these were once definite conquests made...by our ancestors in their attempts to get the chaos of their crude individual experiences into a more shareable and manageable shape. They proved of such sovereign use as *Denkmittel*<sup>44</sup> that they are now a part of the very structure of our mind.  
(James 1909, p. 42)

As the latter part of the first (famous) quote below articulates, the thesis James held is one according to which true propositions are exactly those that prove useful *over the long term*:

*The ‘true’...is only expedient in the way of our thinking, just as ‘the right’ is only the expedient in the way of our behaving.* Expedient in almost any fashion; and expedient in the long run and on the whole of course; for what meets expediently all the experience in sight won’t necessarily meet all farther experiences equally satisfactorily.  
(James 1907, p. 106)

G.E. Moore held that this instrumentalist account of truth is “intensely silly”.<sup>45</sup> There are compelling grounds for agreeing with Moore. For the moment, here are four:

- (1) There is more than one way for a belief to be useful. One way is that the belief leads to the fulfilment of a prediction; another is that it leads to the satisfaction of a desire. But, what if my belief in the weather turning for the worse is a successful prediction but also frustrates my desire to have a picnic in the park?
- (2) Shouldn’t the pragmatist specify *for whom* the belief is intended to be useful for? If the benefactor of the true belief is the agent believing then it might quite conceivably be the case that belief in *p* is useful for Smith while belief in not-*p* is useful for Jones. But this violates the adequacy condition on any theory of truth that it not violate the principle of non-contradiction. Moreover, it commits pragmatism about truth to a form of relativism, something widely regarded as a *reductio* of any theory on the nature of truth. {I expand on

---

<sup>43</sup> James 1909, p. 43. On page 42, James includes such concepts as the notion of time and space as single continuous receptacles; the distinction between thoughts and things, matter and mind; and the separation of fortuitous from regularly caused connections.

<sup>44</sup> Instruments of thought.

<sup>45</sup> Moore 1908 (1959), p. 115.

this type of criticism below}. Indeed, James explicitly commits himself to relativism about truth:

[I]n any concrete account of what is denoted by 'truth' in human life, the word can only be used relatively to some particular trower.<sup>46</sup> Thus, I may hold it true that Shakespeare wrote the plays that bear his name, and may express my opinion to a critic. If the critic both be a pragmatist and a baconian, he will in his capacity of pragmatist see plainly that the workings of my opinion, I being what I am, make it perfectly true for me, while in his capacity of baconian he still believes that Shakespeare never wrote the plays in question.  
(James 1909, p. 146)<sup>47</sup>

We must side with Bertrand Russell in finding this position totally unintelligible. Who wrote *Hamlet* is a "question of fact, totally independent of what anybody now living may think".<sup>48</sup>

- (3) Some false beliefs are useful. The fishing industry still employs a system of navigation based on Ptolemaic astronomy. Though it cannot be denied that true beliefs are on the whole useful and, by and large, useful beliefs are true, there are exceptions. The falsity of my belief that the train's departure time is midday could be useful if the facts conspire to be so; I would not have avoided the train crash that killed so many passengers had I held the true belief that the train was set to depart at 1 pm.
- (4) Even if it so happens that all and only useful beliefs are actually true, the instrumentalist definition of truth has unacceptable modal consequences. As Russell argues, in divorcing truth from the facts, this theory implies that it *might* be useful to believe that A exists even if A does not exist; and it *might* be useless to believe that A exists even if A does exist. Hence, it follows on the pragmatic theory that, absurdly, it might be true that A exists even if A does not exist; and it might be false that A exists even if A does exist:

The [instrumentalist] account of truth assumes, so it seems to me, that no one takes any interest in facts, and that the truth of the proposition that your friend exists is an adequate substitute for the fact of his existence.  
(Russell 1910b (1910a), p. 123)

---

<sup>46</sup> A believer.

<sup>47</sup> This occurs in "Two English Critics" – in James 1909 – and is a response to Bertrand Russell's "Transatlantic Truth" that appeared in the *Albany Review* of January 1908.

<sup>48</sup> Russell 1959, p. 134.

Our pre-theoretic, folk theory of truth, according to which a proposition is true means that the proposition agrees with reality, can schematically be represented thus:

(T) It is true that  $p$  iff  $p$

As there are both actual and possible counterinstances to ‘it useful to believe that  $p$  iff  $p$ ’, (T) fails when ‘useful to believe’ is substituted for ‘true’. Hence, truth is not utility.

## 2.5 *The Non-Viability of Epistemic Theories of Truth*

There is a deep sense in which epistemicism about truth violates common-sense. It does much violence to our ordinary ways of thinking and talking. Others have felt this: opposing the Hempelian view that propositions are comparable *only* with other propositions, Schlick made the following pertinent observation:

I have been accused of maintaining that statements can be compared with facts. I plead guilty. I have maintained this. But I protest against my punishment...I have often compared propositions to facts; so I had no reason to say that it couldn't be done. I found, for instance, in my Baedeker the statement: "This cathedral has two spires." I was able to compare it with 'reality' by looking at the cathedral, and this comparison convinced me that Baedeker's assertion was true.  
(Schlick in Macdonald 1954, p. 232)

Bertrand Russell, too, in fighting the cause of common-sense (here in opposition to Neurath) says:

Take first the necessity of empirical statements about words, e.g., 'Neurath says so-and-so'. How do I know this? By seeing certain black marks on a white ground. But this experience must not, according to Neurath and Hempel, be a ground for my assertion that Neurath says so-and-so. Before I can assert this, I must ascertain the opinion of mankind, and especially of my cultural circle, as to what Neurath says. But how am I to ascertain it? I go round to all the scientists of my culture circle, and say: 'what does Neurath say on p. 364?' In reply I hear certain sounds, but this is an experience, and therefore does not give any ground for an opinion as to what they said. When A answers, I must go round to B, C, D, and the rest of my culture circle, to ascertain what they think A said. And so on throughout an endless regress. If eyes and ears do not enable me to know what Neurath said, no assemblage of scientists, however distinguished, can enable me to know. If Neurath is right, his opinions are not known to me through his writings, but through my decisions and those of my culture circle. If we choose to attribute to him opinions completely different from those which he in fact holds, it will be useless for him to contradict, or to point to pages in his writings; for by such behaviour he will only cause us to have experiences, which are never a ground for statements.  
(Russell 1940, pp. 139-140)

A further violation of common-sense is that epistemicism renders truth relative. There can be no fact of the matter as to what Neurath says, to use Russell's example, so it is possible for what



Neurath says to be true only relative to the “culture circle’s” point of view. Hence, it all depends on the culture milieu you refer to. A commitment to epistemically-fashioned theories of truth compels one to be committed either to *alethic relativism* – ‘the truth value of a statement is relative to a system of beliefs’ – or to *idealism* – ‘how the world is depends upon what humans happen to believe about the world’ or ‘the Earth has only rotated around the Sun since Copernicus’. Either position is incoherent.

Alan Musgrave (see Musgrave 1997 (1999)) has a compelling expression of this anti-epistemicist argument.<sup>49</sup> It consists of first assuming the following: that an adequate theory of truth for any language must yield an instance of the Tarskian *T*-scheme, *S* is true if and only if *P*, for every sentence of that language (suitably constrained so that no contradiction ensues). Combining the *T*-scheme with an epistemic theory of truth commits one to unacceptable, idealistic viewpoints. For example, consider the following epistemic theory of truth:

*S* is true if and only if Arhat thinks that *P*

This ‘theory’, together with the *T*-scheme, entails:

*P* if and only if Arhat thinks that *P*

No-one would be willing to accept that, as a consequence, things in the world are the way I think they are. If so, we could either (a) give up the *T*-scheme, or (b) give up epistemically fashioned theories of truth. Pursuing (a) would force us into an unsustainable position. The scheme is crucial to our common notion of truth; to have a correct grasp of the notion of truth requires that we accept the scheme and all its instances as primitively compelling. Each instance of the *T*-scheme is

---

<sup>49</sup> In arguing for the anti-epistemicist cause, Musgrave also has the following to say: “An immediate consequence [of epistemic truth theories] is relativism about truth. If A’s belief that *P* possesses the internal feature and B’s belief that *P* does not, then *P* is true for A and false for B....

*To avoid relativism (and preserve the laws of truth) subjective truth theorists tend to go ideal and to the long run. For example, coherence theorists will say that something is true (by definition) if it ‘coheres’, not with your or my beliefs, but with the beliefs an ideally rational inquirer would have in the long run.*

*Of course, such moves immediately threaten the anti-sceptical virtues (if virtues they be) of [epistemic] truth theories. What God will be coherently believing at the end of time is just as inaccessible to you or me as is truth-as-correspondence. (I think it is more inaccessible.)*

*To solve these problems, [epistemicists] tend to adopt a policy of flipping back and forth. When epistemological concerns are paramount, they stay subjective; when semantic concerns are paramount, they go ideal and to the long run. The resulting fandango is one of the least edifying sights in philosophy. Let us view it no more.” (Musgrave 1989 (1999), p. 168).*

necessarily, conceptually true; it is rendered true by the concept of truth. How then can there be any room for the epistemic status of a proposition being necessary and sufficient for its truth?<sup>50</sup> So, some might argue, we have no option but to release epistemic theories of truth from the set of theories we are willing to be committed to.<sup>51</sup>

Indeed, epistemic theories of truth may have politically pernicious consequences. The claim that ‘there is no objective truth’ amounts to a denial of the possibility that historical and political claims have an objective truth.<sup>52</sup> The following is a pertinent quote from George Orwell’s *1984*. The quote begins with O’Brien leaning over a torture device to speak to Winston, who is lying on the device:

“Power over matter - external reality, as you would call it - is not important. Already our control over matter is absolute”. For a moment Winston ignored the dial. He made a violent effort to raise himself into a sitting position, and merely succeeded in wrenching his body painfully. “But how can you control matter?” he burst out. “You don’t even control the climate or the law of gravity. And there are disease, pain, death”. O’Brien silenced him by a movement of his hand. “We control matter because we control the mind. Reality is inside the skull. You will learn by degrees, Winston. There is nothing that we could not do. Invisibility, levitation - anything. I could float off this floor like a soap bubble if I wish to. I do not wish to, because the Party does not wish it. You must get rid of those nineteenth-century ideas about the laws of Nature. We make the laws of Nature.”...

“But the world itself is only a speck of dust. And man is tiny, helpless! How long has he been in existence? For millions of years the earth was uninhabited.”

“Nonsense. The earth is as old as we are, no older. How could it be older? Nothing exists except through human consciousness.”

“But the rocks are full of bones of extinct animals, mammoths and mastodons and enormous reptiles which lived here long before man was ever heard of.”

“Have you ever seen these bones, Winston? Of course not. Nineteenth century biologists invented them. Before man there was nothing. After man, if he could come to an end, there would be nothing. Outside man there is nothing.”

“But the whole universe is outside us. Look at the stars! Some of them are a million light-years away. They are out of our reach forever.”

“What are the stars?” said O’Brien indifferently. “They are bits of fire a few kilometres away. We could reach them if we wanted to. Or we could blot them out. The earth is the centre of the universe. The sun and the stars go round it.” Winston made another convulsive movement. This time he did not say anything. O’Brien continued as though answering a spoken objection: “...Do you suppose it is beyond us to produce a dual system of astronomy? The stars can be near or distant, according as we need them. Do

---

<sup>50</sup> William Alston 1996, p. 209 raises this point. Szubka 2003 contains a discussion of this issue.

<sup>51</sup> In Larry Laudan’s view: “*The displacement of the idea that facts and evidence matter by the idea that everything boils down to subjective interests and perspectives is – second only to American political campaigns – the most prominent and pernicious manifestation of anti-intellectualism in our time.*” (Laudan 1990, p. x).

<sup>52</sup> See Deborah E. Lipstadt’s *Denying the Holocaust: The Growing Assault on Truth and Memory*. I thank Jeffrey Ketland for pointing this reference out to me.

you suppose our mathematicians are unequal to that?<sup>53</sup> Have you forgotten doublethink?”  
(Orwell 1949, pp. 277-278)

Epistemic theories of truth equate truth with an epistemic property. In the case of coherentism, a proposition is true just in case it coheres with some specified set of (held) beliefs; for Peirceans, truth is that opinion to which investigators converge in the limit of inquiry; for Jamesians, truth is utility. We have seen compelling reasons to doubt their viability as theories on the nature of truth. This chapter does not pretend to be exhaustive in its coverage of all the forms of epistemicism on offer; there are other types, such as that account given in Putnam 1981 where truth is (idealized) rational acceptability,<sup>54</sup> or in Dummett 1959 where truth is warranted assertibility. But there is a central criticism applicable to all types. They focus on the question ‘On what grounds can I hold *p* true?’ Epistemic theories exhibit a deep confusion about the difference between a theory of knowledge or rationality and a theory of truth, one that is a consequence of not heeding the Russellian maxim of keeping the nature of truth and the criterion of truth distinctly apart. The epistemicist’s focus is what it is for a proposition to be *adopted* or *counted* as true, as opposed to what it is for a proposition to *be* true. Consider, for example, the following telling remark by James:

A new opinion *counts* as ‘true’ just in proportion as it gratifies the individual’s desire to assimilate the novel in his experience to his beliefs in stock.  
(James 1907, p. 36. Emphasis added.)

Epistemic theories of truth are better construed as theories of *rationality*: verifiability, coherence, the potential for leading to successful action etc. are all relevant to rationality. They may be viable contenders for what an adequate theory of rationality looks like, but one must insist that they are not viable contenders for an adequate theory of truth in itself.

---

<sup>53</sup> Over the past two centuries, it has been well known among mathematicians that it is possible to obtain an earth-centred system by coordinate transformation. It would be interesting to know whether Orwell was aware of this.

<sup>54</sup> Here, the circumstances pertaining to the epistemic status of a proposition would be *ideal* just in case all the relevant evidence is readily available (both Alston 1996 p. 194 and Schmitt 1995, p. 113 understand “epistemically ideal conditions” in this way). There is a straightforward objection to this view on truth. There are propositions that are true but *not* justified in epistemically ideal conditions. Consider, for example, Thomas Nagel’s “What Is It Like to Be a Bat?” (1974). It seems reasonable to suppose that there is a fact of the matter concerning what it is like to be bat. Certain propositions concerning the inner life of an echo-locating creature are clearly true. And, yet, we, as cognizers, are cognitively closed off from accessing such facts, no matter what amount of physical evidence we might possibly be able to muster.

# CHAPTER 3

## CORRESPONDENTISM

Per conformitatem intellectus et rei veritas definitur. Unde confirmitatem cognoscere, est cognoscere veritatem.<sup>55</sup>

Thomas Aquinas, *Summa theologiae*

The basic idea underlying the correspondence theory of truth is given by the slogan ‘truth is correspondence with fact’ or ‘truth is agreement with reality’. For many, the central notions of ‘correspondence’ (or ‘agreement’) and of ‘fact’ have been found wanting; most notoriously because allegedly they presuppose the concept of truth itself. It is not entirely clear, it is argued, how the notion of ‘fact’ is to be specified beyond articulating it to be that which makes a proposition true. Also, what *kind* of correspondence is in question here? Arguably, the only kind of understanding involved in saying that true propositions correspond with the world, that they somehow ‘picture’ it, is the metaphorical kind. Surely there is more involved in explicating the idea of correspondence than that involved in the employment of a metaphor? But again, it is not clear that anything can be said about the correspondence relation other than its being the relation which a proposition stands to the world when it is true. Consequently, many have found alternative (and often more attenuated) expressions for what the correspondence theory is, variously attempting to sidestep what they take to be deficiencies in the underlying concepts.

This chapter intends to outline what is right about the correspondence construal. The principal objections levelled against it are, I shall argue, mistaken. I shall also point out how the claim that truth bearers may, without circularity, reflect or correspond to extralinguistic fact has been comprehensively accomplished by the logico-semantical techniques of Alfred Tarski in his celebrated work(s) on truth. Tarski’s explication does not, in other words, fall foul of the common objection (initially due to Frege)<sup>56</sup> that the idea of correspondence necessarily presupposes truth.

### 3.1 *The Primal Attraction of the Correspondence Theory*

We should remind ourselves of the lesson learned from Russell in the previous chapter. In asking after what truth is we are not seeking the supply of conditions that are jointly necessary and

---

<sup>55</sup> “Truth is defined as the agreement of intellect and thing. Hence recognizing that such an agreement obtains is recognizing that something is true”.

<sup>56</sup> See, for example, Frege 1979, pp. 139-140.

sufficient for a claim's being true. What we are after is what the concept *consists in*, that is to say, an *intensional* definition of truth as opposed to a purely extensional one. We want to answer the question: what does it *mean* to say of the proposition that *p*, '*p* is true'? By way of analogy, if you were to list the various positions that manifest checkmate in the game of chess then you would not say that this furnished you with an understanding that to checkmate your opponent is the objective of the game. We need to appreciate the difference between *criterion* and *meaning*; we should not confuse (as the epistemicist does) one for the other when seeking an answer to the question 'What is truth?',<sup>57</sup>

First, we outline what the primal attraction of the correspondence idea is. We also make an initial attempt at understanding the terms 'correspondence' and 'fact', how they fit into what might properly be called a correspondence theory, and address popular objections while keeping a (keen) eye on identifying the reach, and boundaries, of the theory.

The slogan 'truth is correspondence with fact' is widely considered to be not only correct but platitudinous, so much so that it can be regarded as the 'everyman view'. Thus the *Oxford English Dictionary* supplies the following piece of lexicography: "Truth, *n.* Conformity with fact; agreement with reality". What is this common denominator, 'correspondence', involved in our shared understanding of the notion? It seems to amount to the claim that "truth...is a matter of accuracy."<sup>58</sup> Consider, for example, the fact that I want to go on a bus ride tomorrow. I need to buy a ticket for my intended journey, which I do today, and place the ticket in my (only) inside coat pocket. To validate my place on the bus I need to demonstrate to the bus driver that I have bought this ticket. I can only do so by showing the bus driver the ticket. I succeed in doing this by reaching for the ticket in my coat pocket. It is precisely because I believe that *it is true that* the ticket is in this coat pocket that accounts for why reaching for this coat pocket succeeds in my being able to (ultimately) obtain a seat on my intended bus. My belief that the ticket is in my coat pocket, in other words, is *validated* by a corresponding state of affairs.<sup>59</sup> The critic (most notably the deflationist) might respond by saying that the correspondentist is begging the question here. I succeed only

---

<sup>57</sup> Cf. Russell: "If you wish to know whether a certain book is in a library, you consult the catalogue: books mentioned in the catalogue are presumably in the library, books not mentioned in it are presumably not in the library. Thus the catalogue affords a criterion of whether a book is in the library or not. But even supposing the catalogue perfect, it is obvious that when you say the book is in the library you do not mean that it is mentioned in the catalogue. You mean that the actual book is to be found somewhere in the shelves." (Russell 1910b (1999), p. 75).

<sup>58</sup> Searle 1995, p. 200.

<sup>59</sup> *Ceteris paribus*: that is, there is no bus driver strike on the day of my intended bus journey, etc.

because I believe that the ticket in my coat pocket *and (indeed) the ticket is in my coat pocket*.  
Invoking truth is superfluous.

But there is something crucially missing from such a ‘deflating’ account. Compare why this account:

(a) I believe that the ticket is in my coat pocket and the ticket is in my coat pocket

is better at explaining my success than this one:

(b) I believe that the ticket is in my coat pocket and the ticket is on my bedside table

(a) is better because the fact it picks out corresponds to my belief that the ticket is in my coat pocket while the fact picked out in (b) crucially does not so correspond; *it is true* that the ticket is in my coat pocket in that this claim corresponds to the fact that the ticket is in my coat pocket.<sup>60</sup> Had the organization of the world been such that the ticket was on my bedside table then reaching for my coat pocket would have been redundant. *It is impossible to account for preferring (a) to (b) in a fashion that eschews any reference to truth so understood*. Surely, without being able to invoke this kind of explanation for why I succeed in getting the ticket from my coat pocket it would be entirely mysterious as to why I do so succeed.

Many authors have written extensively on the compelling power of the correspondence idea – that the truth of a statement implies that the content of the statement represents or somehow mirrors aspects of the ‘external’ world. Blackburn, for example, writes:

What makes true the belief that there is a cat in the garden, is there being a cat in the garden. There being a cat in the garden is...a very different kind of thing from there being a system of beliefs, controlled in whatever way we like which would have this belief as a member. It is the cat which we respond to, not anything mental. We are good instruments for detecting cats, and that is why, if we are careful, our beliefs about them tend to be true...I am a good signaller of the presence or absence of cats. My belief responds to the cat.

(Blackburn 1984, p. 243)<sup>61</sup>

---

<sup>60</sup> This observation is also made in Vision 2004, p. 2. Vision’s example involves Maud and accounts for her success in fetching the umbrella from the rack arising from the truth of her belief in the umbrella being in the rack (and nothing less).

<sup>61</sup> Also see Kitcher 2002, p. 197 for a defence of this idea.

Our shared folk-theory of truth – that precedes undertaking the burden of any further analytical or explanatory obligations – consists in the commonsensical view that what marks true propositions out from false ones is their being distinguishable by reference to things which are not themselves propositions, but something extralinguistic. Celebrated twentieth century proponents of this idea were Russell<sup>62</sup> and G.E. Moore.<sup>63</sup> Russell’s pupil, Ludwig Wittgenstein, is famous for articulating an early philosophy heavily pregnant with correspondence ideas and notions:

2.21 The picture agrees with reality or not; it is right or wrong, true or false.

4.06 A proposition can be true or false only in virtue of being a picture of reality.  
(Wittgenstein 1922)

But no theory of truth can amount just to the claim: ‘A proposition is true if and only if it corresponds with fact’. This is a *platitude* being “sufficiently bromidic to be acceptable to common sense” as John Dewey once put it.<sup>64</sup> So something more is required for the phrase to work as a philosophical explanation of truth. We need ‘corresponds’ and ‘fact’ to be independently understood, to have their own significance, so that putting them together as ‘corresponds with fact’ succeeds in offering a *genuine* elucidation of the idea, and in particular one capable of a fruitful theoretical development.

### 3.2 The Correspondence Relation

Following Crispin Wright, we might say that one of the things which should be implicit in the phrase ‘agrees with’ when it occurs in expressions like ‘What Dummett said agrees with reality’ or in the phrase ‘corresponds to’ when it occurs in expressions like ‘any statement asserting the

---

<sup>62</sup> See Russell 1912a, pp. 128-129.

<sup>63</sup> See Moore 1953, p. 277. Moore was only an *eventual* adherent of the correspondence view (as was Russell). His initial view, one kindred with the view of ‘good’ he articulated in *Principia Ethica*, was that truth is a simple, unanalyzable, intuitable property. This view was informed principally by his overall rejection of the neo-Hegelian Absolute Idealism dominating the British philosophical scene under the influences of T.H. Green and F.H. Bradley in the late nineteenth century. Moore rallied against idealism’s monism, insisting that reality consists of concepts which combine to form (mind-independent) propositions. Arguing against the Hegelian notion that the unity of a proposition depends on the synthesizing activity of consciousness, Moore came to believe that a true proposition is not one which corresponds with reality, but actually *is a part of* reality.

<sup>64</sup> Dewey 1911, p. 33. If these platitudes were all that needed to be said, then what would the difference between statements that are true and those that are false amount to? We would need a more ‘serious’ attempt in our explication of truth, surely. See Blackburn 1984, p. 225 for an expression of this thought.

existence of unicorns corresponds to no fact' is a *seriously dyadic* relation.<sup>65</sup> To illustrate what this amounts to, to infer:

(a) There is something into which Mukti fell

from:

(b) Mukti fell into the swimming-pool

is to adopt a seriously dyadic attitude to 'fell into' in (b). So, commitment to a correspondence conception of truth results in licensing the inference:

(c) There is something with what Dummett said agrees

from:

(d) What Dummett said agrees with reality

Frege, most notably, disagreed with the relational construal. According to Frege:

Correspondence is a relation. This is contradicted, however, by the use of the word 'true' which is not a relation word and contains no reference to anything else to which something must correspond.  
(Frege 1911 (1999), p. 86)

Was Frege right? Although the predicate '...is true' appears from its grammatical form to be monadic, unlike 'agrees with' or 'corresponds to', it does not follow that no relation is signified by it. Russell has taught us to be wary of being misled by the casual surface structure of natural language. By way of analogy, consider the monadic predicate '...is a spouse', which, not appearing to signifying a relation either, actually means 'There is somebody to whom  $x$  is married'. Spouse-hood signifies a *relational* property. As Russell once put it, "the difference between a true belief and a false belief is like that between a wife and a spinster."<sup>66</sup>

---

<sup>65</sup> See Wright 1992, p. 83.

<sup>66</sup> Russell 1948, p. 165.



### 3.2.1 Deflationism and the Correspondence Relation

We mentioned earlier that ‘a proposition is true if and only if it corresponds with fact’ is platitudinous, acceptable to everyone, and this includes the deflationist. Deflationism, however, is properly a rival view of truth to correspondentism and, as the introductory chapter made clear, holds that – if a property at all – truth is an insubstantial property whose existence owes entirely to the logico-linguistic utility it provides. Despite its being a rival view, deflationism does have features in common with correspondentism – endorsement of the platitude should indicate as much. Importantly for this section, the shared features include endorsement of the idea that truth involves some kind of correspondence relation obtaining. We closely examine deflationism in chapters 4 and 5 but it would be prudent here to explore this connection with the correspondence theory, and to specifically identify in which respects, if at all, correspondentism’s correspondence relation contrasts with that of the deflationists.<sup>67</sup>

Taking our cue from Armour-Garb & Beall, let us putatively first consider the deflationism-correspondentism contrast lying in what kind of property each position takes truth to be: for the deflationist truth is an insubstantial property, for the correspondentist it is a substantial (relational) property.<sup>68</sup> However, clearly such a characterization is simply too imprecise to be of much use.<sup>69</sup> For if we were to consider the *T*-schema:

*S* is true if and only if *p*

where ‘*S*’ is replaced by a name of the sentence on the right-hand side of the equivalence, then, according to correspondentists, its instances express a correspondence between the sentence and the state of affairs represented on the right. For example, in examining the following instance:

‘Snow is white’ is true if and only if snow is white

---

<sup>67</sup> Chapter 4, section 1, deals further with the contrasting features of both these positions.

<sup>68</sup> This is in their introduction to the anthology *Deflationary Truth* (Armour-Garb, B. & Beall, J.C., 2005a). Both attempt to identify what is at stake first by considering that as substantialists view truth to *be* a property, the deflationary contrast would be the denial that truth has any property-like features. But, as Armour-Garb & Beall correctly recognize, this cannot be the contrasting feature as there are deflationists who think that truth *is* a property – Paul Horwich being the notable example.

<sup>69</sup> Cf. Armour-Garb, B. & Beall, J.C., 2005a, p. 2.

a correspondentist would say that this states a correspondence between the sentence ‘Snow is white’ and the state of affairs described on the right: truth is being defined in terms of an attribution of various non-linguistic ‘worldly’ properties to various non-linguistic ‘worldly’ entities.<sup>70</sup> But, of course, deflationists appeal to exactly the *same* schema in advancing their insubstantiality claim, viz. truth has no underlying nature whose task it is for philosophers to unpack. In their view, any sentence is equivalent (logically, or cognitively, or necessarily) to its truth predicated counterpart and this is what the *T*-schema conveys.<sup>71</sup> What, then, distinguishes correspondentism from deflationism?<sup>72</sup> We take up Jan Woleński’s invitation to mark a distinction between *weak* and *strong* correspondentism. On this distinction, Jan Woleński has written:

Under the strong concept, the relation of correspondence is seen as a similarity...of truth-bearers and reality, but under the weak one, this relation consists in asserting a semantic correlation between expressions and extralinguistic entities; the relation of designation (reference, denotation) is an example of semantic correlation.  
(Woleński (web article), p. 16)<sup>73</sup>

George Pitcher made a similar distinction, one between *correspondence-as-congruence* and *correspondence-as-correlation* (cf. Pitcher 1964, p. 10). Strong correspondence involves a general characterization of the structural relationship between sentences and worldly states of affairs, while weak correspondence involves truth just being dependent on how the world is.<sup>74</sup> Correspondentism is to be understood as advocating ‘correspondence’ in the *strong* sense, the sense of correspondence-as-congruence. Many deflationists are happy to advance the idea that the truth of a proposition ‘depends’ on extralinguistic matters. It is just that this dependence is not a perfect match nor, indeed, does it imply anything substantial about truth. Here are a couple of telling comments from arch deflationists:

---

<sup>70</sup> Cf. Sher 1999, pp. 134-135.

<sup>71</sup> See Horwich 1990, pp. 7-8, Quine 1960, p. 24, Ayer 1936 (1971), p. 116 for examples. This part of the deflationary proposal on truth is discussed in the following chapter.

<sup>72</sup> Douglas Patterson 2003 and Panu Raatikainen’s 2007 make similar observations.

<sup>73</sup> Woleński’s first articulated this distinction in 1989, in the paper (jointly written with Peter Simons) “De Veritate: Austro-Polish Contributions to the Theory of Truth from Brentano to Tarski”. Interestingly, they see Aristotle as espousing both types of view in his works. The famous “*To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.*” in *Metaphysics*, Γ, 7, 1011<sup>b</sup>25 conveys a weak correspondence theory of truth, whilst, by contrast, a strong version is to be understood from “*He who thinks the separated to be separated and the combined to be combined has the truth, while he whose thought is in a state contrary to that of the objects is in error.*” in *Metaphysics*, Γ, 7, 1051<sup>b</sup>3.

<sup>74</sup> In *Truth*, A.R. White (1970) identifies this distinction as one between ‘correspondence *with*’ and ‘correspondence *to*’. An illustration of the correspondence-as-congruence view is when a key *corresponds with* its keyhole, while an example of the correspondence-as-correlation view is an entry in a ledger that *corresponds to* a sale.

No sentence is true but reality makes it so. The sentence ‘Snow is white’ is true...if and only if *real* snow is *really* white.

(Quine 1970, p. 10. Emphasis added.)

...minimalism [Horwich’s particular brand of deflationism]...does not deny that truths *do* correspond—in *some* sense—to the facts; it acknowledges that statements owe their truth to the nature of reality...

(Horwich 1998a, p. 104)

So deflationists should be understood as correspondentists about truth, but only in the *weak*, correspondence-as-correlation sense.<sup>75</sup>

We pause here to rehearse what we have learnt so far. In doing justice to our folk theory of truth – encapsulated in slogans like ‘truth is correspondence with fact’ – the correspondence theory (contrary to what Frege thought possible) views the property expressed by ‘true’ as analyzable in terms of the relation expressed by ‘corresponds to’.<sup>76</sup> This relation is seriously dyadic and understood in the strong, correspondence-as-congruence sense;<sup>77</sup> this contrasts with deflationism which, though also committed to the folk theory encapsulated in the *T*-schema, observes a correlative relation alone obtaining between truth bearers and worldly facts. We aim to demonstrate precisely what a strong correspondence theory looks like, but first we turn to an examination of the notion of ‘fact’.

### 3.3 What are Facts?

There appear to be undeniable cases of concrete facts. Among them, we might include the facts that John F. Kennedy was assassinated on 22<sup>nd</sup> November 1963, that London is in England, the fact that Paris is the capital of France, and that the philosophical study of truth is taxing. Moreover, the fact that Paris is the capital of France appears to be a different fact from the fact that London is in England. However, in their role as securing what the truth of a statement consists in, it is not clear

---

<sup>75</sup> Notice how even this weak sense of ‘correspondence’ is sufficient to exclude epistemic theories of truth if one finds the type of correspondence expressed by instances of the *T*-scheme undeniable (which one should).

<sup>76</sup> As in: *x* is true iff there is a fact *y* such that *x* corresponds to *y*. Importantly, it cannot be the case that the correspondentist views the monadic property expressed by ‘true’ as *identified* by the binary relation expressed by ‘corresponds to’.

<sup>77</sup> Indeed, this implies that some traditional correspondentists must be taken to advocate only a weak position, i.e. a position not strong enough to allow us to see what is tangibly at issue between correspondentism and deflationism. J.L. Austin’s view on truth would be such an example. He wrote: “A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions (the one to which it “refers”) is of a type with which the sentence used in making it is correlated by the descriptive conventions.” (Austin 1950 (1999), pp. 152-153).

how facts are to be individuated independently of the statements whose truth they are playing a part in constituting. Typical constructions involving the fact-operator '*it is a fact that...*' say exactly the same thing as typical constructions involving the truth-operator '*it is true that...*': 'it is a fact that John F. Kennedy was assassinated on 22<sup>nd</sup> November 1963' is only a stylistic variant of 'it is true that John F. Kennedy was assassinated on 22<sup>nd</sup> November 1963' and nothing more; it might appear then that 'fact' has been created only to furnish us with a perfect substitute for 'true statement'. Fact-talk – so this typical metaphysical objection goes – is entirely parasitic on truth-talk; one is paraphrasable in terms of the other. Hence, facts have been mistakenly reified, whence the charge that they are not fit to serve as truthmakers; they have no independent resources of their own with which to confer truth to statements. This sentiment is widely shared:

We could...eliminate the word 'fact' from the language altogether and substitute for it the longer expression 'true proposition'...[i]n ordinary usage they are surely identical, in that whenever we assert that something is a fact we could assert that it is true without change of meaning.  
(Woozley 1966, p. 12)

I contend that "a fact" and "a true proposition" mean identically the same thing. The evidence is that the expression "It is true that..." can always be substituted for "It is a fact that..." and vice-versa; and that nothing can be deduced from what either expression formulates, which can not be deduced from what the other formulates.  
(Ducasse 1940, p. 710)

By nominating facts as determiners of truth, are we not just fabricating substance, projecting "from true sentences for the sake of correspondence"?<sup>78</sup> The convergence of fact-talk with truth-talk can be explained in one of two ways, as Vision explains:

- (a) Facts are nothing more than reifications of a style of speaking.
  - (b) Since it is a, perhaps *the*, chief job of statements to state facts, it is understandable that the same linguistic resources would be marshalled to formulate both.
- (Vision 2004, p. 66)

Despite the sound observation that the language in which fact-talk is couched is (roughly) equivalent to that in which truth-talk is, it does not follow that the correspondentist is magically attempting to prise rabbits<sup>79</sup> out of her metaphysical hat. While it is a trite observation that facts are specified by means of statements to which they correspond, it is premature to take the ontological category to be illusory once a structural similarity between a style of speaking and the ontological category has been observed. The (equally) viable option of the linguistic structure *actually*

---

<sup>78</sup> Taken from Quine 1987, p. 213.

<sup>79</sup> Fact-bits?

reflecting in some interesting way a genuine ontological category is not considered.<sup>80</sup> Why could the statement that ‘snow is white’, as one kind of thing, not correspond to the fact that snow is white, as another kind of thing? It is irresponsible for opponents of facts to embrace (a) without at the same time dismissing (b).<sup>81</sup>

### 3.3.1 *The Slingshot Argument*<sup>82</sup>

Opponents of facts-in-themselves may still have a case, however, and we now turn to what appears to be a very powerful argument that the very notion of fact is incoherent. The slingshot argument is a reductio purporting to show that if there are facts at all there is only one to which all true statements correspond. If facts are not non-trivially individuable then this presumably must render the notion of fact, and by implication the correspondence theory, incoherent. This would indeed be a reductio ad absurdum.

The slingshot was first developed by Alonzo Church in the introduction to Church 1956 (pp. 24-25). He considers the following set of sentences:

- (1) Sir Walter Scott is the author of *Waverley*
- (2) Sir Walter Scott is the man who wrote twenty-nine *Waverley* novels altogether
- (3) The number, such that Sir Walter Scott is the man who wrote that many *Waverley* novels altogether, is twenty-nine
- (4) The number of counties in Utah is twenty-nine<sup>83</sup>

Given that the name or description ‘the author of *Waverley*’ is replaced by another (‘the man who wrote twenty-nine *Waverley* novels altogether’) which has the same reference, i.e. Scott, (1) and (2)

---

<sup>80</sup> By way of comparison, Vision observes (2004, p. 67) that the words ‘violets on a mahogany table’ are used to talk about violets on a mahogany table *and* to talk about the mirror image of violets on a mahogany table; however, we do not infer that the mirror image is not different from the violets on the mahogany table.

<sup>81</sup> Another issue: how can true statements enter into causal relations in the way that facts do? It is not the true statement ‘the British rate of interest was set at the German level’ that caused *Black Wednesday* in 1992 but the fact that the British rate of interest was set at the German level that caused *Black Wednesday* in 1992.

<sup>82</sup> Barwise & Perry 1981 first coined the name “slingshot argument”.

<sup>83</sup> A formal presentation of the slingshot was presented in Church 1943, pp. 299-301, where he criticised Carnap’s idea that sentences stand for, or designate, propositions. Church demonstrated that if propositions exist then there can be at most only the one proposition.

must have the same reference.<sup>84</sup> The same applies to (3) and (4): the latter is obtained from the former by replacing the description ‘the number, such that Sir Walter Scott is the man who wrote that many *Waverley* novels altogether’ by another referring to the same object (the number twenty-nine). Given that (2) and (3) are “if not synonymous...[then] at least so nearly so as to ensure its having the same denotation” for Church, they too must have the same reference. Therefore, (1) and (2), (2) and (3), and (3) and (4) have the same reference when taken pairwise, which means that (1) and (4) must do too. So, (1) and (4) have differing senses *yet* the same reference. The only semantic feature that they retain is their truth-value.<sup>85</sup> Church used this example to demonstrate that sentences with non-equivalent senses may still have equivalent referents, and that no matter what the reference is each will have the same one.

Davidson lost no time in using the argument explicitly against the correspondence theory. And this was in spite of his earlier self who ardently defended both the correspondence theory and the thesis that Tarski delivered such a theory. Davidson said of his earlier self: “[t]here is...serious reason to regret having said that a Tarski-style truth theory was a form of correspondence theory” (Davidson 1990, p. 304) identifying the slingshot argument explicitly as “the real objection to correspondence theories” (*Ibid.* p. 303).<sup>86</sup> Davidson’s later work is considered by many to present an influential and thorough critique of interpreting Tarski’s legacy in terms of correspondence. We develop Tarski’s theory later and demonstrate how precisely Davidson was wrong to defect from the position his earlier self occupied. First, however, let us examine Davidson’s version of the slingshot argument. In “True to the Facts”, Davidson first considers when

(S) The statement that *p* corresponds to the fact that *q*.

---

<sup>84</sup> Gödel 1944 preferred ‘signification’ to ‘reference’ when translating the German ‘Bedeutung’ as, in his view, it conveys the ordinary meaning of the term better. In footnote 4 of Gödel 1944 (1964), p. 213, he says, “*I use the term ‘signify’ ...because it corresponds to the German word ‘bedeuten’ which Frege...used in this connection*”. Church 1956 referred to it as ‘denotation’.

<sup>85</sup> Church took the argument to elucidate the Fregean idea that the *Bedeutung* of a sentence is that sentence’s truth-value. Although Frege never explicitly gives a slingshot argument against the correspondence theory, Church argued that implicit in Frege’s view was that all true sentences refer to just one thing. According to Frege 1892, one of the things that linguistic competence requires is that, as far as the meaning of a singular term is concerned, one needs to be sensitive to the distinction between its *Bedeutung* and its *Sinn* (or sense). Given the principle according to which the replacement of a singular term embedded in a sentence by another having the same reference leaves the reference of the whole unchanged (an application of the ‘principle of compositionality’), it is possible to construct sentences that differ in sense but not in reference.

<sup>86</sup> There should be no misunderstanding of Davidson on this point. Lorenz Krüger (Krüger 1995) and Ilkka Niiniluoto (2004), to give just two examples from among plenty, both confirm that Davidson here cites the slingshot argument as the reason for his defecting from the position his earlier self occupied.

would be true.<sup>87</sup> Clearly (S) is true when both  $p$  and  $q$  are replaced by the same sentence. However, unless facts are to be understood as mere reflections of true sentences, there ought to be true instances of (S) where  $p$  and  $q$  are not identical. Davidson then observes that since (as an example) Naples satisfies the following description ‘the largest city within thirty miles of Ischia’, then the statement that Naples is farther north than Red Bluff corresponds to the fact that Red Bluff is farther south than the largest city within thirty miles of Ischia. Given further that Naples *also* satisfies the description ‘the largest city within thirty miles of Ischia, and such that London is in England’, then “we begin to suspect that if a statement corresponds to one fact, it corresponds to all.”<sup>88</sup> This suspicion is validated as long as the following two principles are assumed to hold:

- The statements replacing ‘ $p$ ’ and ‘ $q$ ’ are logically equivalent
- ‘ $p$ ’ and ‘ $q$ ’ differ only in that a singular term has been replaced by a co-extensive singular term

For Davidson then, the argument is this:

Let ‘ $s$ ’ abbreviate some true sentence. Then surely the statement that  $s$  corresponds to the fact that  $s$ . But we may substitute for the second ‘ $s$ ’ the logically equivalent ‘(the  $x$  such that  $x$  is identical with Diogenes and  $s$ ) is identical with (the  $x$  such that  $x$  is identical with Diogenes)’. Applying the principle that we may substitute coextensive singular terms, we can substitute ‘ $t$ ’ for ‘ $s$ ’ in the last quoted sentence, provided ‘ $t$ ’ is true. Finally, reversing the first step we conclude that the statement that  $s$  corresponds to the fact that  $t$ , where ‘ $s$ ’ and ‘ $t$ ’ are any true sentences.<sup>89</sup>  
(Davidson 1969 (1984), p. 42)

Formally, the argument looks like this {where ‘ $(\iota x)$ ’ means ‘the  $x$  such that... $x$ ...’ }:

- |   |  |
|---|--|
| 1. $s$  | Premise  |
| 2. $(\iota x)(x = d \wedge s) = (\iota x)(x = d)$ | From 1., given substitution <i>salva veritate</i> of logical equivalents |
| 3. $(\iota x)(x = d \wedge t) = (\iota x)(x = d)$ | From 2., given substitution <i>salva veritate</i> of co-referring terms  |
| 4. $t$  | From 3., given substitution <i>salva veritate</i> of logical equivalents |

---

<sup>87</sup> Davidson 1969 (1984), p. 41. Davidson presented a similar argument in 1967a (2001), pp. 18-20.

<sup>88</sup> Davidson 1969 (1984), p. 42.

<sup>89</sup> Given that the mechanics of the argument rest on minimal logical machinery and appears to cripple a venerable philosophical thesis in one fell swoop, one sees why the David and Goliath metaphor has been employed.

All four lines of this argument correspond to the same fact. In “The Structure and Content of Truth”, Davidson argued that the moral to draw from this is that it:

...trivialize[s] the concept of correspondence completely; there is no interest in the relation of correspondence if there is only one thing to which to correspond, since, as in any such case, the relation may well be collapsed into a simple property: thus, “s corresponds to the universe”, like “s corresponds to (or names) the True”, or “s corresponds to the facts” can less misleadingly be read “s is true”. (Davidson 1990, p. 303)<sup>90</sup>

In fact, as Gödel indicated in “Russell’s Mathematical Logic”, the slingshot being loaded here can be made even more powerful.<sup>91</sup> Gödel employed a notion of equivalence weaker than that of logical equivalence – what Stephen Neale has termed *Gödelian equivalence*<sup>92</sup> – one obtaining between sentences like ‘Fa’ and ‘a = (∃x)((x = a) ∧ Fx)’<sup>93</sup>: they are to “mean the same thing”.<sup>94</sup> This is weaker because true identity statements remaining true under the substitution of logically equivalent

---

<sup>90</sup> In fact, Davidson thought that this demonstration of facts collapsing into ‘the Eleatic one’ (Gödel’s terminology) has repercussions that go beyond pinpointing the vacuity of correspondentism – we ought to banish representational semantics altogether: “If [the slingshot argument] is right, and I am convinced it is, we ought also to question the popular assumption that sentences, or their spoken tokens, or sentence-like entities or configurations in our brains, can properly be called “representations”, since there is nothing for them to represent. If we give up facts as entities that make sentences true, we ought to give up representations at the same time, for the legitimacy of each depends on the legitimacy of the other”. (Davidson 1990, p. 304).

Convinced by the slingshot and its significance, Davidson in his later work came to defend a primitivism about truth, a position he occupied while lamenting the fact that: “[w]e are still under the spell of the Socratic idea that we must keep asking for the essence of an idea, a significant analysis in other terms, an answer to the question what makes this an act of piety, what makes this, or any, utterance, sentence, belief, or proposition true. We still fall for the freshman fallacy that demands that we define our terms as a prelude to saying anything further with or about them.” (Davidson 1996 (1999), p. 320). The importance of the concept of truth, according to Davidson’s later philosophy, forecloses on the possibility of finding a foundation for it reaching deeper into the bedrock (cf. Davidson 1996 (1999), p. 309). Contrary to Tarski who believed that ‘true’ is explicitly definable, Davidson expressly treated ‘true’ as a primitive predicate in his later works.

<sup>91</sup> See Gödel 1944 (1964), pp. 213-214. His whole discussion of this issue appears on pp. 213-215 in the 1964 edition of the Benacerraf & Putnam anthology *Philosophy of Mathematics*. Gödel did not present a slingshot argument, but indicated what one would look like.

<sup>92</sup> See Neale 1995, 1997 & 2001. The accounts defended in these are considered by many to be authoritative commentaries on slingshot arguments (Graham Oppy 1997, 2004 being his most prominent rival). In them, Neale reconstructs the argument along Quinean lines as a means by which to demonstrate what logical features certain general sentential connectives must have in order for them to be non-extensional. And this has a bearing on the issue of the viability of such intensional entities as facts.

<sup>93</sup> Gödel also assumed that any sentence standing for a fact can be rephrased into a predicate-argument form. This is assumption (2) in Gödel 1944 (1964), p. 214, footnote 5: “every proposition [that] “speaks about something”...can be brought to the form  $\phi(a)$ .” Clearly, without this assumption the slingshot envisaged will only hold for all true *atomic* sentences.

<sup>94</sup> Gödel failed to expand on how he intended “mean the same thing” to be taken here, but given his way of parsing signification in the main text it seems reasonable to suppose that he intended it to mean ‘signify the same thing’ in that ‘Fa’ and ‘a = (∃x)((x = a) ∧ Fx)’ stand for the same fact. So they preserve subject matter, adding no new non-logical vocabulary (cf. Neale 2001, p. 131).



statements entails *Gödelian equivalence*, but not vice-versa. The argument runs through assuming this weaker equivalence principle.

There remains a question, however. The validity of the slingshot argument presumably depends on the validity of the semantics of definite descriptions adopted. Such a semantics, it is argued, must – in the case of the Church-Davidson version – (i) render ‘*s*’ and ‘ $\iota x(x = d \wedge s) = \iota x(x = d)$ ’ logically equivalent, (ii) declare the definite descriptions ‘ $\iota x(x = d \wedge s)$ ’ and ‘ $\iota x(x = d \wedge t)$ ’ co-referential when ‘*s*’ and ‘*t*’ are true, and (iii) treat definite descriptions as singular, referring terms.<sup>95</sup> And similarly for the Gödelian version: if one wished to hold that definite descriptions are singular terms that refer and that sentences standing for facts are determined by the referents of their component parts then one cannot hold that ‘*Fa*’ is somehow a different fact from the fact that ‘ $a = (\iota x)((x = a) \wedge Fx)$ ’. However, as many have pointed out, on Russell’s theory of descriptions,<sup>96</sup> definite descriptions do not stand for objects, or refer to things; they are not referential because they are not singular terms. According to this theory, any sentence of the form ‘the *F* is *G*’ ought rather to be understood as belonging to the class of quantificational-predicational expressions; they are on a par with the quantifiers ‘every’, ‘some’, ‘a’, ‘no’ which are syncategorematic terms that, when combined with nominal expressions, yield noun phrases (cf. Russell 1905 (1956), p. 42).<sup>97</sup> Thus, the sentence ‘the *F* is *G*’ is equivalent to the corresponding sentence ‘there is one and only one *F*, and it is *G*’, formalized as ‘ $\exists x(Fx \wedge \forall y(Fy \rightarrow y = x) \wedge Gx)$ ’,<sup>98</sup> giving us the wherewithal to turn any sentence containing definite descriptions into an equivalent that is description-free.<sup>99</sup> Given this

---

<sup>95</sup> Smullyan 1948, p. 39 was the first to spot the significance of treating definite descriptions as names (or singular terms), but in the context of the collapsing argument for modality found in Quine’s work. García-Carpintero & Pérez Otero 1998 highlight its significance for Davidson’s argument against facts too.

<sup>96</sup> See Russell 1905, 1918.

<sup>97</sup> This employed the method of *contextual definition* that precipitated the so-called ‘context principle’ of Frege’s: that it is only in the context of a sentence that a word has meaning. This is why Russell took denoting phrases to be ‘incomplete’ (cf. Whitehead & Russell 1925, p. 67). According to Quine, contextual definitions are the reward “for recognising that the unit of communication is the sentence not the word.” (Quine 1981, p. 75).

<sup>98</sup> This is essentially a variant of Principia Mathematica \*14.01:  $[\iota x\phi] \Sigma(\iota x\phi) =_{df} \exists x(\forall y(\phi(y) \equiv y=x) \wedge \Sigma(x))$ , where  $\Sigma(\iota x\phi)$  is a sentence containing  $\iota x\phi$  and  $\Sigma(x)$  is the result of replacing any occurrences of  $\iota x\phi$  by  $x$  in  $\Sigma(\iota x\phi)$ ; and  $[\iota x\phi]$  indicates that the description has minimal scope. This is the proposition defining definite descriptions. \*14.02 is often also taken to be involved but essentially it just deals with talk of existence (statements of the form ‘the *F* exists’) and is included to deal with the fact that Russell regards genuine singular terms as never failing to refer, so existence cannot be represented by a predicate symbol.

<sup>99</sup> Russell’s theory, therefore, furnishes us with the means by which to make expressions such as ‘the present King of France does not exist’ meaningful *without it having to refer*. It is important to note that (a) the domain of quantification in Russell’s description-free paraphrases must be understood as contextually indicated – by saying ‘the cat is on the mat’ clearly I do not mean to say ‘there is one and only one cat in the whole universe and it is on the mat’, and (b) Russell’s theory is intended to apply only to definite descriptions that involve singular reference to particulars – so objecting that his paraphrases fail to deal with such locutions as ‘the cats are on the mat’ or ‘the port in Oxford colleges’ is irrelevant.

theory, and the ‘principle of compositionality’,<sup>100</sup> it cannot then be the case that both ‘ $a = (\lambda x)((x = a) \wedge Fx)$ ’ and ‘ $a = (\lambda x)((x = a) \wedge a \neq b)$ ’ are obtainable from each other from the substitution of co-referring terms. And so, it does not follow that they stand for the same fact. The property of being  $F$  is part of the fact corresponding to ‘ $a = (\lambda x)((x = a) \wedge Fx)$ ’ but not the fact corresponding to ‘ $a = (\lambda x)((x = a) \wedge a \neq b)$ ’. Indeed, ‘ $Fa$ ’ has a truthmaker that is an entirely different (singular) fact from the general fact making ‘ $a = (\lambda x)((x = a) \wedge Fx)$ ’ true, and so a Russellian need not accept that they stand for the same fact. On Russell’s theory, ‘ $a = (\lambda x)((x = a) \wedge Fx)$ ’ is shorthand for ‘ $\exists y[(y = a) \wedge Fy] \wedge \forall z(((z = a) \wedge Fz) \rightarrow (z = y)) \wedge (a = y)$ ’,<sup>101</sup> and so its truthmaker –  $a$  – need not be the truthmaker of ‘ $Fa$ ’.

Gödel, however, was hesitant to endorse this application of Russell’s semantics,<sup>102</sup> saying:

...I cannot help feeling that the problem raised by Frege’s puzzling conclusion [that all true sentences have the same signification] has only been evaded by Russell’s theory of descriptions and that there is something behind it which is not yet completely understood.  
(Gödel 1944 (1964), p. 215)

Gödel was right to be hesitant. Excluding definite descriptions from the primitive notation just creates the illusion of a solution, since, as Church showed,<sup>103</sup> the argument can be reformulated in terms of set-abstraction operators where there is no question that they refer (in the standard model, whose existence we can reasonably assume here).<sup>104</sup> To argue, as Neale does, that Gödel’s slingshot “forces philosophers to say something about the semantics of definite descriptions...as soon as they

<sup>100</sup> Katz & Fodor first coined the phrase “the principle of compositionality” in 1963. There they state that this principle is highly intuitive, not to be questioned; indeed, Gödel himself viewed it as an “*obvious axiom*” (Gödel 1944 (1964), p. 450).

<sup>101</sup> Strictly speaking, Russell’s theory that definite descriptions are ‘incomplete’ means that they have to be analyzed *within a sentence*; they are not themselves sentences (nor equivalent to sentences).

<sup>102</sup> Indeed, as did Quine (albeit in a different context): “*Looking upon quantification as fundamental and constant singular terms as contextually defined, one must indeed concede the inconclusiveness of a criterion of referential opacity that rests on interchanges of constant singular terms.*” (Quine 1953b (1966), p. 174).

<sup>103</sup> Church 1943.

<sup>104</sup> This is what Davidson’s slingshot looks like when the *iota* operators are replaced by set abstracts: let  $s$  and  $t$  abbreviate true sentences. The following then is a *valid* argument, with each line corresponding to the same fact:

1. $s$	Premise
2. $\{x: x = d \wedge s\} = \{x: x = d\}$	From 1., given substitution <i>salva veritate</i> of logical equivalents
3. $\{x: x = d \wedge t\} = \{x: x = d\}$	From 2., given substitution <i>salva veritate</i> of co-referring terms
4. $t$	From 3., given substitution <i>salva veritate</i> of logical equivalents

posit entities to which sentences are meant to correspond”<sup>105</sup> would, therefore, be wrong-headed: too much weight has been placed on imaginary problems concerning the *iota* operator.<sup>106</sup>

Does the slingshot argument force us to revise our ordinary speech permitting us to speak, as it appears to be doing, of nothing more than one fact? What the argument does demonstrate is that the folk theory of facts has (quite possibly) unacceptable consequences, and rather graphically exhibits one. It shows that there is a need to move from ‘folk’ fact-based semantics, which doesn’t work properly, to a more scientific semantics in elucidating the idea of correspondence. The conceptual apparatus provided by Tarski, it will be argued, succeeds where the ‘folk’ fact-talk failed: facts are a very important, almost indispensable, ontological category whose intensional structure is actually preserved in a theory like Tarski’s and it is owing to this kind of faithfulness to the structure of facts which serves to deflect the slingshot. But there is still clarificatory groundwork yet to be done on the nature of facts.

### 3.3.2 *Is Correspondentism the View of Metaphysical Realism?*

Evidently, the relata involved in the correspondence relation are required to be distinct. In particular, the truth-maker is a non-mental, non-linguistic, ex-cogitational entity. It is owing to this that some see correspondentism as entailing realism.<sup>107</sup> However, this entailment is easily seen not to be valid. In what follows I shall follow Michael Devitt and take realism to be the doctrine that tokens of (most, current) common-sense and scientific physical types objectively exist independently of the mental.<sup>108</sup> There are thus two distinctive elements to this doctrine: an *existence* element concerning what entities, and what types of entity (physical and, presumably, mental) exist, and an *independence* element concerning their independent nature. Correspondentism does require the reality responsible for the truth of sentences to be objective (so it entails the independence element) but it is entirely silent on the existence element. Correspondentism does not tell us which entities exist for it does not tell us which sentences are true; the theory tells us only what the grounds are for using the concept of truth by telling us what it is for a sentence to be true.

---

<sup>105</sup> Neale 2001, p. 223.

<sup>106</sup> This does not mean that the friend of facts has nothing instructive to say; only that she is not forced into making any commitments about the *semantics* of descriptions (cf. Rodriguez-Pereya 2003).

<sup>107</sup> See, for example, Putnam 1981, p. 273, and Ellis 1985, pp. 50-51.

<sup>108</sup> Devitt 1991, p. 23.

### 3.3.3 *Taking Stock*

What has the preceding discussion told us? The correspondence theory of truth states that the truth of a proposition consists in some relation of correspondence to a truth-maker, or ‘fact’, which may or may not be cognizable. But what elevates it above triviality is that it understands the word ‘correspondence’ as a real relational property; a seriously dyadic one. Contrastively with the deflationary relation of correspondence-as-correlation, the correspondentist advocates seeing the correspondence relation in the strong, correspondence-as-congruence sense. The truth-maker is understood to be a different type of entity from the truth-bearer – the latter is a linguistic item while the former is an extralinguistic one. However, the slingshot argument arguably shows that a semantics based on a folk theory of facts, i.e. one in which the inner structure of facts are not explicitly identified and subjected to compositional reasoning, is unsuited to participating in a rigorous theory of correspondentist truth. It has however been claimed that there is a fatal objection to any such correspondence theory, an objection to which I now turn.

### 3.4 *Ramsey’s Ladder*

Consider, for example, the judgement that there is a cat on the mat. Check whether this judgement corresponds to the facts (by inspecting the mat, etc.); confront the judgement with reality. But there is only the one standpoint here – that of the inspector – and not another constituted by having the judgement and the facts before us so that we can check whether the one corresponds with the other. What would such a confrontation look like?<sup>109</sup> Deflationists (such as Frank Ramsey who first voiced the issue) want to argue that since there is no difference between judging a situation from judging that we are judging that situation, the latter is essentially absorbed into the former. For Ramsey the interplay of attitude and content is crucial here. He considered the following three statements:<sup>110</sup>

- (a) The earth is round
- (b) It is true that the earth is round
- (c) Anyone who believes that the earth is round believes truly

He remarked that these are all equivalent, and continued:

---

<sup>109</sup> Cf. Rorty 1986 (1999), p. 327.

<sup>110</sup> Ramsey 1991.

...we can think of the earth being round as a possible subject of belief and say “If you think the earth is round, you think truly” and this amounts to no more than that the earth has the quality you think it has when you think it is round, i.e. that the earth is round.  
(Ramsey 1991 (2001), p. 441)

If the judgement that  $p$  contains exactly the same content as it is true that  $p$  and for it is really true that  $p$ , and similarly for it is really and truly a fact that it is true that  $p$ , and so on, then surely every rung on this apparent ladder of ascending and increasing insight cannot be considered more informative than the very bottom rung. Blackburn dubs this “Ramsey’s Ladder”, cautioning philosophers who attempt to climb it and profess a clearer philosophical view ‘from the top’.<sup>111</sup>

In their introduction to the *Oxford Readings in Philosophy* anthology on truth, Blackburn and Simmons give the following description of Ramsey’s Ladder:

Thinking this way, it may appear, involves thinking of two distinct standpoints. There is the standpoint we occupy when we judge that  $p$ . Then in addition there is the standpoint we occupy when we step back, and judge that the judgement that  $p$  indeed bears the right relation to the fact that  $p$ . But then it is reasonable to urge that there are not two distinct standpoints here, but only the one... There are indeed mental processes that we can call ‘standing back’: becoming cautious about  $p$ , checking one more time whether  $p$ , and so on. But these are all processes of reflecting and checking whether  $p$ . None of them introduce a separate topic, and yet the correspondence theory seems to demand that there is both a separate topic and a separate standpoint from which it can be judged. It is here that we see the illusory nature of the idea that we can step outside our own skins or stand on our own shoulders.  
(Blackburn & Simmons 1999, p. 7)

Why should it be considered unintelligible to have a distinctly alternative standpoint or attitude from that occupied when judging that  $p$ , in which the judgement that  $p$  plays a part? Why can one not judge  $p$  and judge one’s comparison of this judgement with the fact that  $p$ ? Let’s call the first kind of judgement a ‘first-order’ judgement and the second kind of judgment a ‘second-order’ judgement. For the Ramsey objection to hold, it would require that the second-order judgement *just is* the first-order judgement as it is not possible for the episode of (having the attitude of) first-order judging to play two separate roles at the same time: that of (having the attitude of) second-order judging *and* being part of the content of this same second-order judgement. It is, therefore, not possible to have a second-order attitude to  $p$  *unless the content of this attitude is crucially altered so as to manifest two distinct standpoints with respect to  $p$* . And this is what is required for there to be a legitimate, distinct, second standpoint. But it is difficult to see why this requirement is supposed

---

<sup>111</sup> Blackburn 1998, p. 78 and pp. 294-298. Frege posed a challenge that in some ways seems rather similar (see, for example, Frege 1911 (1999), p. 87), but the lesson each wishes to draw differs. Frege thought that a correspondence account involves a logical regress, one according to which truth is necessarily presupposed. Ramsey thought that a correspondence account involved a hierarchy of standpoints which is equally untenable.

to hold and, moreover, why second-order judgements just cannot be different from their first-order counterparts. There is, for example, nothing unintelligible about noting the fact that while I can see this cat on the mat, I cannot see myself seeing this cat on the mat; but this does not incline me to think that the content of ‘me seeing this cat on the mat’ is not different from the content (or object) of my observation – namely, this cat on the mat. My inability to see this cat on the mat and see myself doing this in the very same episode does not support the idea that there is no content to my seeing this cat on the mat; if this is so, then why could one not give content to judging something about the relation between one’s judgement that  $p$  to what it is a judgement about – namely,  $p$  – in a fashion that intelligibly differs from simply making the judgement that  $p$ .

### 3.5 Dr. Tarski, or: How Truth Theorists Learned to Stop Worrying and Love Semantics

We saw earlier that correspondentism sees the truth-concept as a real relational property (in the strong, correspondence-as-congruence sense) and that it is in this that we are to identify what the notion consists. Part of Tarski’s achievement was that he showed truth to be equivalent to *satisfaction*.<sup>112</sup> This is why Karl Popper was so delighted in Tarski’s work. In *Objective Knowledge*, Popper says the following about his pre-Tarskian concerns with truth:

It was not so much the antimony of the liar which frightened me, but the difficulty of explaining the correspondence theory: what could the correspondence of a statement to the facts be?  
(Popper 1972, p. 320)

Tarski, then, had come to Popper’s rescue:

[w]hat can we possibly mean if we say that truth is correspondence with the facts (or with reality)?...Tarski solved...this apparently hopeless problem by making use of the semantic metalanguage, reducing the idea of correspondence to that of ‘satisfaction’ or ‘fulfilment’.  
(Popper 1959, note \*1)<sup>113</sup>

And so, accordingly, Popper thought:

---

<sup>112</sup> If we make the restrictive assumption that all individuals in the domain concerned are named then introducing ‘satisfaction’ as a separate item is clearly not necessary.

<sup>113</sup> Indeed, despite what he was later to defend, in his earlier thoughts on truth Davidson appreciates this point: “*The semantic concept of truth as developed by Tarski deserves to be called a correspondence theory because of the part played by the concept of satisfaction; for clearly what has been done is that the property of being true has been explained, and non-trivially, in terms of a relation between language and something else.*” (Davidson 1969 (1984), p. 48)

Tarski's greatest achievement...is that he rehabilitated the correspondence theory of...truth which had become suspect. He vindicated the free use of the intuitive idea of truth as correspondence to the facts. (Popper 1960, p. 223)<sup>114</sup>

Popper overlooks further reasons for applauding Tarski – one being Tarski's forestalling Frege's famous regress objection. Frege believed that any investigation into determining what the truth of a proposition consists in must presuppose that truth is involved in that very determination exercise.<sup>115</sup> To determine whether 'snow is white' is true, the correspondentist must assay whether or not the claim that snow is white corresponds to the relevant chunk of reality; but then she must also determine whether or not *it is true* that such a correspondence obtains, and whether or not it is true that *this* is true, and so on. Tarski's vindication of correspondentism (crucially) involves explicating correspondence in the sense above in a manner that does not presuppose the notion itself (something that deeply concerned Frege about reductive analyses of truth). There is the further reason for applause: the slingshot conclusion is not derivable in Tarski's theory. Fact-talk, and the correspondence theory, became unfashionable among some philosophers after the attacks of Frege on the one hand, and of the slingshot on the other. This section hopes to succeed in showing that both these objections are successfully avoided by the homomorphism account implicit in a Tarskian truth-definition.

We now recount Tarski's approach to truth, laying to rest some popular concerns as we go along, and then attend to what are taken to be fundamental objections to Tarski's work.

### 3.5.1 Correspondentist Aims

Tarski explicitly intended his conception of truth "to catch hold of the actual meaning of an old notion", a notion he found expressed in the "classical" (Aristotelian) conception and according to which truth is correspondence with reality.<sup>116</sup> Thus, contrary to a claim sometimes expressed in the

---

<sup>114</sup> Part of the suspicion referred to by Popper here is that in the heyday of logical positivism when Tarski was doing his research the concept of truth was viewed as inherently metaphysical and, thus, philosophically suspicious. Only if the notion could be rendered scientifically respectable (which most commentators understand Tarski to have interpreted as *physical*, especially since Hartry Field's 1972 paper "Tarski's Theory of Truth") could it then be accepted as meaningful. There is, of course, some disagreement on this issue. Frost-Arnold 2004, for example, argues *against* the idea that Tarski was motivated by physicalism. Among other themes, Frost-Arnold mentions that Tarski appreciated the prejudices of his audience in the 1930s and needed to placate them enough to be able to convey his overall package on truth.

<sup>115</sup> See, for example, Frege 1979, pp. 139-140.

<sup>116</sup> Cf. Tarski 1944 (1999), p. 116. See also Tarski 1936a (1983), p. 155. Tarski borrowed the expression "*classical conception of truth*" from Kotarbiński 1929, a reference Tarski often consulted. The tradition of the 'Lvov-Warsaw

literature, Tarski did intend his conception to be a theory of correspondence.<sup>117</sup> Tarski called his conception of truth ‘semantic’ and understood the discipline of semantics as dealing with the relation between linguistic entities and the (extralinguistic) entities they are about, or referred to (*cf.* Tarski 1944 (1999), p. 120).

### 3.5.2 *Truth is Relational*

Tarski 1944 noted Frege’s claim that in contradistinction to the relational concepts of designation and definition the word ‘true’ is of a different logical nature,<sup>118</sup> since unlike them it appears to be monadic. But it does not follow that truth is not best understood as a relational concept. Tarski argues that it is *because* truth is semantical – picking out a relation between the expressions of a language and what they refer to – that it designates a certain class of sentences. He writes:

A characteristic feature of the semantical concepts is that they give expression to certain relations between the expressions of language and the objects about which these expressions speak, or that by means of such relations they characterize certain classes of expressions.  
(Tarski 1936a (1983), p. 252)

Though not explicitly articulating a relation between sentences and the objects they are about, truth does characterize certain classes of expressions in a way that is *determined* by such relations.<sup>119</sup> And so, truth is relational.<sup>120</sup>

### 3.5.3 *Satisfaction*

Tarski introduced the notion of satisfaction in ‘Der Wahrheitsbegriff in den Formalisierten Sprachen’ (1936a). The basic idea can be illustrated thus:

Sequence of objects  $s = \{s_1, s_2, s_3, \dots\}$  satisfies (-in- $L$ ) the formula ‘ $x_i < x_j$ ’ iff  $s_i$  is less than  $s_j$ .

---

School’ in which Tarski operated is extensively documented by Jan Woleński in Woleński 1989a (with P. Simons), 1989b and (web article). Also see K. Szaniawski 1989 and A. Rojszszczak 1998, 2002.

<sup>117</sup> For example, Susan Haack says the following: “Tarski does not regard himself as giving a version of the correspondence theory.” (Haack 1978, p. 114). See also Read 1995, p. 22.

<sup>118</sup> Tarski 1944 (1999), p. 120.

<sup>119</sup> *Cf. Ibid.*, p. 121.

<sup>120</sup> This is especially so given that “the simplest and the most natural way of obtaining an exact definition of truth is one which involves the use of the semantic notions, e.g. the notion of satisfaction.” (*Ibid.*)



The point of having a satisfaction relation can be adumbrated as follows. Suppose we have the following:

- (i) an object language  $L$  and a metalanguage  $ML$ , with  $ML$  containing  $L$ , and
- (ii) for each sentence  $\phi$  in  $L$ , a name  $\ulcorner \phi \urcorner$  in  $ML$ , and
- (iii) a *translation function* from  $L$  to  $ML$ ,  $\Xi(\cdot)$ , mapping every  $L$ -formula to an associated  $ML$ -formula i.e., if  $\phi$  is an  $L$ -formula then  $\Xi(\phi)$  is the associated  $ML$ -formula.

If we let  $MT$  be the metatheory containing a predicate intended to mean ‘true-in- $L$ ’ and let  $Tr$  be this predicate then, following Tarski, we stipulate the following definition for a materially adequate truth theory:

$Tr$  is a truth predicate for  $L$  within a metatheory  $MT$  in  $ML$  just in case,  
for each  $L$ -sentence  $\phi$ , every  $ML$ -sentence  $Tr(\ulcorner \phi \urcorner) \leftrightarrow \Xi(\phi)$  is a theorem of  $MT$ <sup>121</sup>

If the object language under consideration contains a countable number of sentences then “the general definition has to be, in a certain sense, a *logical conjunction* of all the partial definitions.”<sup>122</sup> This is unproblematic for a correct definition of truth can be supplied by completing the following disjunctive scheme (under the usual constraints):

$x \in Tr$  if and only if either  $x = x_1$  and  $p_1$ , or  $x = x_2$  and  $p_2$ , or  $x = x_3$  and  $p_3$ ,  
or  $x = x_4$  and  $p_4$ , or  $x = x_5$  and  $p_5$ , or  $x = x_6$  and  $p_6$ , ..., or  $x = x_n$  and  $p_n$ ...

and this is materially adequate. However, any interesting formalized language has a potential infinitude of sentences for which such a (finite) list-like definition cannot be compiled. Allowing for logical complexity by introducing the truth-functional logical connectives to operate on sentences of  $L$  guarantees this. Accordingly, “the idea of using the recursive method suggests itself”<sup>123</sup> so that the truth of any complex (molecular) formula is determined by the truth of the atoms that compose it. For instance, the recursive clause for ‘disjunction’ operating on  $L$  would be:

---

<sup>121</sup> I present material adequacy this way because of the significance Tarski attaches to the translatability of object language sentences into the metalanguage (see Tarski 1936a (1983), p. 187).

<sup>122</sup> Tarski 1944 (1999), p. 120. Emphasis added.

<sup>123</sup> Tarski 1936a (1983), p. 189.

For any sentences  $S_1$  and  $S_2$ ,  $S_1 \vee S_2$  is true-in- $L$  iff  $S_1$  is true-in- $L$  or  $S_2$  is true-in- $L$

But this presupposes that *all* object language sentences are truth-functional. For quantificational languages the above ‘direct’ recursive method would founder; there are logically complex sentences whose atoms are not necessarily closed – what the truth of sentences like  $\forall x F(x)$  consists in cannot be explained in terms of the truth of what its parts consist in as there is no instruction so far as to what the truth of the open sentence  $F(x)$  consists in. Tarski’s solution was to employ satisfaction or, more precisely, *satisfaction by a sequence of objects*.<sup>124</sup>

[I]n general composite sentences are in no way compounds of simple *sentences*...In view of this fact no method can be given which would enable us to define [truth] directly by recursive means. The possibility suggests itself, however, of introducing a more general concept which is applicable to any sentential function, can be recursively defined, and, when applied to sentences, leads us directly to the concept of truth. These requirements are met by the notion of the *satisfaction of a given sentential function by given objects*...

(Tarski 1936a (1983), p. 189)

When the object language  $L$  contains quantifiers, the concept of satisfaction (of a formula relative to a sequence of objects in the domain) is first defined by recursion on the syntactical complexity of formulas. The concept of truth (of a sentence) is then defined as satisfaction by one (and, hence, every) sequence of objects in the domain. A formal presentation of these ideas is provided in the appendix, section 3.7 (ii).

Note how the satisfaction relation can indeed be seen to be a “correspondence relation”, one holding between objects in the world and (in particular, open) formulas. It is certainly a *seriously dyadic* relation for it is needed in order for a theory of truth to be compositional: if one wants the truth of a statement of the form:

*for all  $x$   $F(x)$*

or:

---

<sup>124</sup> The finite subset of objects to which the variables of a complex formula refer are the only relevant satisfiers. So, for instance, the predicate “loves( $x_i$ ,  $x_j$ )” is satisfied by a sequence  $s$  just in case the  $i$ th element of  $s$  loves the  $j$ th element. An alternative way of defining the truth value of complex sentence is the game-theoretical approach (see Hintikka 1996). In Hintikka’s treatment, complex sentences are replaced by a game with the structure of a tree whose branches represent the various ways in which the sentences may be verified or falsified. Truth simply means the existence of a winning strategy for the verifier. Note that game-theoretic semantics gives an objectual treatment of atomic sentences and quantifiers so that, as with Tarski’s theory, the game-theoretic approach can be regarded as a form of the correspondence theory.

*there is an  $x$ , such that  $F(x)$*

to depend semantically on their proper syntactic part:

$F(x)$

in some way then one needs a semantic concept that one can apply sensibly to an open formula rather than just to a closed one (i.e., a sentence). Without satisfaction one would have to fall back on substitutional quantification, where instead of quantified sentences being made true by objects in the world they are made true by their substitution instances in the language, and this should be avoided.<sup>125</sup> Many deflationists, however, would be willing to accept the satisfaction scheme:

*For all  $x$ :  $Sat(x, \ulcorner F(y) \urcorner) \leftrightarrow F(x)$*

Hartry Field, for instance, endorses this scheme in many of his publications.<sup>126</sup> So we have yet to develop seeing how Tarski's theory does justice to a correspondence theory in the strong, *correspondence-as-congruence*, sense.

### 3.5.4 *Truth-in-a-Model*

Tarski's 'Der Wahrheitsbegriff in den Formalisierten Sprachen' (1936a) falls within the early period of his work in semantics and deals explicitly with the "absolute concept of truth" (1983, p. 199). This notion of truth *simpliciter* is not relativized to any model, it is understood to mean the same as correctness in "the class of all individuals" (1983, p. 207), i.e. according to Tarski this one-world domain is assumed to consist of all objects, including individuals, classes and higher-order entities; it is a single and fixed comprehensive universe.<sup>127</sup> But it is the general model-theoretic concept which builds on this and which defines truth relative to any arbitrary domain (introduced by

---

<sup>125</sup> As the argument against the use of substitutional quantification is highly relevant to my discussion of deflationism I delay expanding on this point to Chapter 4.

<sup>126</sup> For example, see Field 1998.

<sup>127</sup> As Woleński and Simons 1989a make clear, Tarski's preoccupation with the absolute conception of truth was a heritage from the Lvov-Warsaw school, influenced in particular by Twardowski, Kotarbiński (to whom Tarski dedicated *Logic, Semantics and Metamathematics*) and Leśniewski, and by their support for the classical Aristotelian definition of truth as correspondence with fact.

Tarski during his mature work in the 1950s)<sup>128</sup> which is of interest. It demonstrates how, relative to whatever convention we happen to select in linking language up with the world (i.e., the interpretation function), a notion of truth in the strong, *correspondence-as-congruence* sense is deliverable. We develop truth-in-a-model in this section.

In his mature work, Tarski adopted the convention according to which the non-logical constants of a language  $L$  (individual constants, one-place predicates, two-place predicates, etc.) are enumerated in a fixed order and their interpretations in a relational structure are then given in the same order. So, for a given interpretation function  $I$  from  $L$  to a domain of individuals  $X$ , we can treat  $(L, I)$  as an interpreted language and define a structure for this language as a *relational system* of the form  $(X, s_1, \dots, A_1, \dots, R_1, \dots)$  with designated elements  $s$ , subsets  $A$  and relations  $R$  that are the  $I$ -images of the vocabulary of  $L$  in the domain  $X$  (i.e. the function  $I$  maps individual constants to elements of  $X$ , predicates to subsets of  $X$  and relations on  $X$ , and function symbols to functions in  $X$ ). Tarski did not make the interpretation function  $I$  explicit but it is clear that such a function from  $L$  to the relational system is presupposed.<sup>129</sup> In textbooks on model theory, it has become standard to explicitly express the link between language  $L$  and a set-theoretic structure  $\mathbf{M}$  by the interpretation function  $I$ .<sup>130</sup> An  $L$ -structure  $\mathbf{M}$  is thus defined as the pair  $(X, I)$ . Language  $L$  is then an uninterpreted syntactic language which becomes interpreted via  $\mathbf{M}$ . For convenience, we follow the standard modern formulation. For atomic sentences of  $L$ , truth in structure  $\mathbf{M}=(X, I)$  is defined by the following conditions:

$$\mathbf{M} \models P(a_1) \text{ iff } I(a_1) \in I(P)$$

---

<sup>128</sup> See Tarski 1954-55 and Tarski & Vaught 1957.

<sup>129</sup> This is arguably confirmed by the fact that in his *Foundations of Logic and Mathematics* and *Introduction to Semantics*, Carnap's own characterization of Tarski's approach (of which he was a great admirer) a *designation* function **Des** corresponding to the pair  $(L, I)$  is explicitly given in Carnap's semantical system  $S$ . For Carnap, **Des** is a language-world relation where individual constants designate individual objects and predicates designate properties. **Des** is first defined for individuals and predicates and then by recursion for sentences. For example, if  $a$  designates snow  $\{\mathbf{DesInd}('a', \text{snow})\}$  and  $P$  designates the property of being white  $\{\mathbf{DesAttr}('P', \text{the property of being white})\}$ , then  $P(a)$  designates the proposition that snow is white  $\{\mathbf{DesProp}('P(a)', \text{snow is white})\}$ . Truth of sentences in semantical system  $S$  is defined as follows:

(C) Sentence  $s$  is true in  $S$  iff there is a proposition  $p$  such that  $s$  designates  $p$  in  $S$  and  $p$

(C) clearly resembles Tarski's  $T$ -schema but does have the advantage of making the semantic connection between sentences and their truth conditions explicit by the relation of designation. Niiniluoto says: "[I]n this respect...the treatment of semantics by Carnap in the late 1930s and early 1940s was more satisfactory than Tarski's (1944) own explanations." (Niiniluoto 1999a, p. 96). While this is certainly true, it should not be overlooked that Tarski was explicit in restricting his attention to *interpreted* languages, i.e. languages assumed to be interpreted in the domain of all objects (see Tarski 1936a (1983a), pp. 166-167 and Tarski 1969, p. 68 ).

<sup>130</sup> See, for example, Chang & Keisler 1973 or Monk 1976.

$$\mathbf{M} \models Q(a_1, a_2) \text{ iff } \langle I(a_1), I(a_2) \rangle \in I(Q)$$

An open formula  $A$  of  $L$  with free variable  $x_i$  is assigned a truth value in the structure  $\mathbf{M}$  by some element of  $X$ . Let  $s = (s_1, s_2, s_3, \dots)$  be an infinite sequence of objects from  $X$ . Then sequence  $s$  satisfies formula  $A$  in structure  $\mathbf{M}$ , i.e. the relation  $\mathbf{M} \models_s A$ , is defined by recursion on the complexity of  $A$ . For example:

$$\mathbf{M} \models_s A \vee B \text{ iff } \mathbf{M} \models_s A \text{ or } \mathbf{M} \models_s B$$

$$\mathbf{M} \models_s \forall x_i A \text{ iff } \mathbf{M} \models_{s(i/b)} A \text{ for all } b \in X$$

where  $s(i/b)$  is the sequence obtained from  $s$  by replacing the  $i$ -th element of  $s$  with  $b$ . The basic clauses for atomic formulas take the following form:

$$\mathbf{M} \models_s P(x_i) \text{ iff } s_i \in I(P)$$

$$\mathbf{M} \models_s Q(a_i, x_j) \text{ iff } \langle I(a_i), s_j \rangle \in I(Q)$$

When  $A$  does not contain occurrences of free variables, i.e. it is a sentence of  $L$ , then it is satisfied in  $\mathbf{M}$  by one sequence  $s$  if and only if it is satisfied by all sequences. Hence, we can define  $\mathbf{M}$ 's being a model of  $A$ , i.e.  $A$  is true in  $\mathbf{M}$ , thus:

$$\mathbf{M} \models A \text{ iff } \mathbf{M} \models_s A \text{ for every } s$$

The later Davidson worried that the ‘correspondences’ found in the recursive clauses for satisfaction were simply artefacts of the axioms necessary to produce the recursions. But the recursive axioms can quite sensibly be justified on independent grounds, in particular on the grounds that truth is itself fundamentally a compositional concept; indeed, the very notion of compositional semantics is often itself identified with that of recursive semantics (see Janssen 1997).

The key components of the above are:<sup>131</sup>

- (i)  $L$  is syntactically well determined, i.e. the  $L$ -sentences have a unique compositional structure<sup>132</sup>

---

<sup>131</sup> I follow Niiniluoto 2002.

- (ii) Truth is defined for sentences of  $L$
- (iii) Truth is defined relative to a structure  $\mathbf{M}$
- (iv) An interpretation function  $I$  mapping the syntactical elements of  $L$  to structure  $\mathbf{M}$  is presupposed
- (v) The relation of satisfaction between open formulas and sequences of objects is presupposed in the recursive definition of truth for quantified sentences<sup>133</sup>

According to the model theoretic approach pioneered by Tarski, truth in a model is relative to an  $L$ -structure  $\mathbf{M}$ .<sup>134</sup> The truth or falsity of an  $L$ -sentence depends on the structure of the world relative to the descriptive vocabulary of  $L$ ; true sentences are therefore not in a relation of correspondence with a non-structured reality but rather to a structure consisting of objects with properties and relations. There are as many  $L$ -structures representing the world as there are languages  $L$  and the same language  $L$  can be interpreted in different domains or by different interpretation functions on the same domain; one might say that the  $L$ -structures represent the *possible worlds* distinguishable within the  $L$ -vocabulary.<sup>135</sup> Relative to whatever convention we choose to link language with the world, Tarski's definition gives us a notion of truth in the strong, *correspondence-as-congruence* sense. In the case of the 'correspondences' found in the recursive clauses for satisfaction, the later Davidson worried that these were simply artefacts of the axioms necessary to produce the recursions.

We have observed that a model (or structure) stands in a satisfaction, or making-true, relation to the sentences of a language when true. Facts have not yet been defined, that is entities to which single true sentences are said to correspond to have not yet been cast into open view. It is still the case that if two distinct sentences of one language are satisfied in a model by its making-true relation then it is that *one and same* model satisfying both the two distinct sentences, but from a correspondence theory in the strong sense one would naturally demand that two distinct sentences are satisfied by

---

<sup>132</sup> This point is emphasized in Woleński 1993 and, as Tarski points out in many places (for instance in Tarski 1944), this requirement is at least approximately satisfied by portions of natural language. So it is incorrect to remark, as Haack does, that Tarski was "*thoroughly sceptical about the applicability of his theory to natural languages*" (Haack 1978, p. 120).

<sup>133</sup> As Niiniluoto emphasizes, except for (iii) and (iv), all these ideas are in fact clearly presented in Tarski 1936a.

<sup>134</sup> And has proven to be successfully applicable not only to artificial formal languages but also to *fragments of scientific, and of natural, languages*. For examples, see Przelecki 1969, Tuomela 1973, Pearce 1987 and Niiniluoto 1987.

<sup>135</sup> So we may say that states of affairs are relative to the languages used to describe them. But it does not follow that truth is therefore in some sense 'epistemic'. For it is up to the world to decide whether, having fixed the meanings of 'London', 'Edinburgh' and 'is south of' for a given language  $L$ , London is south of Edinburgh.

two distinct entities, *viz.* two distinct facts, when true, and by nothing else. I address this in 3.5.5, providing a theory much like Tarski's, fundamentally building on it, in which explicit reference to facts is made.

The model theoretic approach to truth is an extremely flexible tool with which to express relations between  $L$ -sentences and set-theoretic  $L$ -structures. In the case of mathematics, arithmetic say, Tarski's definition allows us to investigate the various structures that are models of such axiomatic theories. If the theory has a unique intended interpretation, like the structure of natural numbers  $\mathbb{N}$  for arithmetic, then truth in  $\mathbf{M}$  captures the standard notion of truth for the language of arithmetic. If what we are particularly interested in is the actual world, then as the actual world is a member of the set of possible worlds, truth in the actual world is a special case of truth in possible worlds.<sup>136</sup> Thus, let  $L$  be a fragment of natural language that speaks about existing objects. Among the  $L$ -structures  $\mathbf{M}$  there is a unique  $L$ -structure  $\mathbf{M}^* = (X^*, I^*)$  where  $X^*$  is the domain of existing objects and  $I^*$  is the interpretation function mapping the syntactical entities of  $L$  to their extensions in  $X^*$ . A sentence of  $L$  is then actually true if it is true in  $\mathbf{M}^*$ .<sup>137, 138</sup>

Despite Wilfrid Hodges's "disconcerting experience" not having found the notion of 'structure', and hence 'truth in a structure' in Tarski 1936a,<sup>139</sup> the early account can be seen as a special case of the model-theoretic approach outlined here.<sup>140</sup> According to Hodges, the central notion of an uninterpreted constant symbol acquiring an interpretation by being applied to a particular structure is completely missing. However, this ignores the fact that Tarski's classic paper follows the Lvov-Warsaw tradition of explicitly restricting attention to *interpreted* languages, i.e. languages assumed to be interpreted in a domain (*cf.* Tarski 1936a (1983a), pp. 166-167). In his early work, in providing his definition of the absolute concept of truth, Tarski assumed the domain interpreting the

---

<sup>136</sup> From the explanations given in Tarski 1944 it is clear that in his earlier work Tarski attempted to define truth in the actual world.

<sup>137</sup> Sentences that are actually true, but not logically true, are then *factually* or *materially* true.

<sup>138</sup> For formal languages, the interpretation function  $I$  is fixed by stipulation. But terms in natural languages have historically determined meanings which are based on conventions. It is owing to these conventions, or language games played by linguistic communities, that statements like "'Russell' refers to Bertrand Russell" express contingent semantical facts. Hence, though this model theoretic account presupposes an interpreted object language, this account is compatible with various methods of fixing reference.

<sup>139</sup> Hodges 1986.

<sup>140</sup> Niiniluoto expresses this view in a number of his publications – see, for example, Niiniluoto 2002. Niiniluoto reminds us also that Tarski often referred back to his 1936 paper for a definition of truth in a structure. Indeed, Vaught 1974 states that Tarski's major contribution in 1936 was to show how ' $\sigma$  is true in  $\mathbf{M}$ ' can be defined inside ordinary mathematics (in *ZF* set theory for example).

language to consist of *all objects*. However, with reference to “the Göttingen school grouped around Hilbert”, Tarski also points out the *relative* notion of “correct or true sentence in an individual domain *a*” (p. 199) according to which “we restrict the extension of the individuals considered to a given class *a*” (p. 199) – where *a* is a subdomain of the domain of all objects. This interpretation is confirmed by Tarski’s remark that such specific individual domains are relevant in connection with the various special sciences (p. 239). Niiniluoto 2004 points out that Tarski’s early work can be understood as occupying a middle position (one world-many languages) between what Hintikka 1997 refers to as ‘language as a universal medium’ (one world-one language) and ‘language as a calculus’ (many worlds-many languages). Indeed, in Tarski 1936b, Tarski complains that previous attempts at supplying a semantics for language “proceeded as though there was only one language in the world” (1983, p. 402).<sup>141</sup> The final move to the calculus view of the general model theoretic approach is taken in the possible worlds semantics which drops the one-domain assumption.

### 3.5.5 Tarskian Truth and Facts

We have seen how in Tarski’s model theoretic account the truth maker is a relational, set-theoretic structure. Tarski avoided speaking about ‘facts’ as such – a consequence, as is well documented, of his following his teacher Tadeusz Kotarbiński who advocated an ontological doctrine rejecting all species of abstract entities such as facts, states of affairs, properties and sets. In his rejection of sets, however, Tarski did not follow Kotarbiński. Tarski is after all renowned for sophisticated mathematical work in set theory. Indeed, Tarski’s own explanation of the concept of satisfaction in ‘Der Wahrheitsbegriff in den Formalisierten Sprachen’, exemplified by:

For every individual *a*, *a* satisfies the sentential function ‘*x* is white’ iff *a* is white<sup>142</sup>

is a condition guaranteeing that the class of objects satisfying the condition ‘*x* is white’ is simply the class of white objects. If the interpretation of the predicate ‘white’ is fixed so that its extension is the set of white objects, then the truth condition for the sentence ‘snow is white’ is formulated as follows:

---

<sup>141</sup> This point is overlooked in objections that claim that the application of model theory to the semantics of natural languages presupposes some absolute conception of the world – one that is fixed and ready, and where the answers to all possible questions is decided (*cf.* Sundholm 1994). Model theory connects each language *L* to a family of *L*-structures, of which one would be a fragment of the real world. This particular structure would only be able to answer those questions that can be asked within the language *L*.

<sup>142</sup> Tarski 1936a (1983), p. 190.



Snow belongs to the class of white things

and it is natural to regard this as a set-theoretic formulation of *the fact that snow is white*.

One might still complain that what a correspondence theory requires is explicit quantification over or reference to facts (or to obtaining states of affairs) and that this is entirely absent in Tarski's work. In other words, this complaint amounts to requiring more from Tarski than simply a scientific precisification of the idea that a proposition is true if and only if it corresponds (congruently) with fact. It is certainly true that in both the early and late stages of Tarski's work truth does not *explicitly* involve any seriously dyadic relation between propositions and facts – there is no primitive, or defined, binary predicate in the language of Tarski's theory intended to apply to pairs of propositions and facts.<sup>143</sup> However, it is perfectly possible to reformulate Tarski's theory – in which states of affairs get *described* – to one in which states of affairs get *referred to*, or *quantified over*. One can demonstrate this in the following way:<sup>144</sup> consider a first-order language  $L$  in which, for simplicity, all individual terms  $t$  are individual variables. A semantic value mapping  $v$ , mapping  $L$ -formulas (also open  $L$ -formulas) to states of affairs can be inductively defined thus:

$$v(P(t_1, \dots, t_n)) = \langle v(P), v(t_1), \dots, v(t_n) \rangle$$

$$v(\text{not } A) = \langle \text{NEG}, v(A) \rangle$$

$$v((A \text{ or } B)) = \langle v(A), \text{OR}, v(B) \rangle$$

$$v(\text{Exists } x_i F(x_i)) = \langle \text{EXISTS\_}x_i, v(F(x_i)) \rangle$$

where NEG, OR, and EXISTS\_ $x_i$  are distinct constituents of states of affairs (perhaps identical to the truth functions that correspond to the logical signs in question, where in the case of the existential quantifier the relevant variable would have to be carried along in some way), where “<...>” are delimiters giving states of affairs hierarchical structure or ‘depth’,<sup>145</sup> and where  $v$  applied to a predicate of arity  $k$  yields a relation of arity  $k$ . Note that this way of setting the

---

<sup>143</sup> This does not mean that one cannot still accept Tarski's theory as a correspondence theory.

<sup>144</sup> We will stick to Tarski's original 1936 theory.

<sup>145</sup> Barwise & Etchemendy 1987 identify the *rhs* of each equality above with set-theoretic tuples. It is not necessary for our purposes to be committed to set theoretic tuples here.

apparatus up means that variables themselves, as opposed to their values, would become components of states of affairs; for otherwise quantification would not be handled correctly as one might see by applying the recursive recipe above to an existentially quantified sentence.<sup>146</sup> Truth can now be defined as follows:

$$Tr(A) \leftrightarrow v(A) \text{ obtains (with respect to every sequence } s \text{ of individuals)}^{147}$$

where “obtains with respect to a sequence  $s$  of individuals” is an undefined, primitive term and whose meaning can be described as follows:

a state of affairs of the form  $\langle \text{NEG}, st \rangle$  obtains with respect to a sequence  $s$  of individuals if and only if  $st$  does not obtain with respect to  $s$

a state of affairs of the form  $\langle st_1, \text{OR}, st_2 \rangle$  obtains with respect to a sequence  $s$  of individuals if and only if  $st_1$  obtains with respect to  $s$  or  $st_2$  obtains with respect to  $s$

a state of affairs of the form  $\langle \text{EXISTS\_}x_i, st \rangle$  obtains with respect to  $s$  if and only if  $st$  obtains with respect to at least one sequence  $s'$  that differ from  $s$  at most at  $i$

where  $st, st_1, st_2$  are states of affairs. In true Tarskian spirit, obtaining is thus relativized to sequences of individuals, just as satisfaction is relativized to sequences. Moreover, such a correspondence theory would include the following as axioms for atomic states of affairs on the one hand and atomic formulas  $P(t_1, \dots, t_n)$  on the other:

a state of affairs of the form  $\langle v(P), x_{k_1}, \dots, x_{k_n} \rangle$  obtains with respect to a sequence  $s$  of individuals if and only if  $v(P)(s(k_1), \dots, s(k_n))$ <sup>148</sup>

---

<sup>146</sup> If one is not happy with variables as components of states of affairs then there are other ways of handling quantification. This approach is advantageous as it stays close to Tarski.

<sup>147</sup> Where the sequence  $s$  is rendered obsolete in the obtaining of  $A$  where  $A$  is a sentence, not an open formula.

<sup>148</sup> In mathematical language, the correspondence-as-congruence view is called a *homomorphism* (Sher 2001 makes this observation on p. 202). This is the Boolean algebraic analogue of an interpretation. Closely related to (and not to be confused with) a homomorphism is an isomorphism – literally the ‘same form’ – which is a special case of a homomorphism. Two objects are isomorphic when they share the same form. But, when one object ‘reflects’ or is an ‘image’ of another object, then the reflection is a *homomorphic image* of its subject, and the above demonstrates how true sentences are homomorphic images of facts. Formally (following Dunn & Hardegree 2001, p. 15), suppose  $L$  is an object language with  $\mathbf{A}(L)$  its associated algebra, and suppose that  $\mathbf{P}$  is an algebra of propositions for this language (to be understood as equivalent to an ‘algebra’ of states of affairs, and which must be appropriate for  $L$  in the sense of their

The Davidsonian point alluded to above – that the recursive clauses for satisfaction factitiously service correspondence since that machinery is only present to enable the recursions – can be addressed here as it was above (Davidson’s point would of course run here *mutatis mutandis* for the recursive clauses for obtaining). There is no need to be so pessimistic since the recursive axioms can quite sensibly be justified on independent grounds, in particular on the grounds that truth is itself fundamentally a compositional concept.

One might wonder where the correspondence relation has disappeared. The answer is simple: it has not disappeared. Use of the semantic value mapping  $v$  from above amounts to correspondence in the following sense:

$$\text{Corr}(A, st) \leftrightarrow v(A) = st$$

where  $A$  is a sentence and  $st$  is a state of affairs. Such a theory is a correspondence theory of truth in a ‘narrow’ sense, *viz.* a proposition is true just in case the proposition (congruently) corresponds to a fact (or to an obtaining state of affairs), and where this very equivalence is spelled out such that facts are quantified over and/or referred to by the theory. It stays very close to Tarski’s original theory; indeed it should be viewed as nothing more than a reformulation or modification of Tarski’s original theory – in particular, the clauses for “obtains with respect to” correspond to the Tarskian clauses for satisfaction, and it is provably materially adequate in the sense of Tarski. It meets the objection from negative facts, disjunctive facts, and so on, and notice also how this is an account of modelling states of affairs which preserves their intensional structure, thus immunizing it from the slingshot objection, i.e. one cannot derive the consequence that only two states of affairs obtain.<sup>149</sup> *Contra* Davidson, we have a model for a theory of facts which makes no allowance for collapsing,

---

existing a one-to-one correspondence between the syntactic operations and their semantic counterparts). Then an interpretation of  $L$  in  $\mathbf{P}$  will be any homomorphism from  $\mathbf{A}(L)$  into  $\mathbf{P}$ , i.e. there is a function  $\mathfrak{t}$  such that the following condition is satisfied by every sentential connective  $c$  in  $L$ :

$$\mathfrak{t}(o_c(\phi_1, \dots, \phi_k)) = O_c(\mathfrak{t}(\phi_1), \dots, \mathfrak{t}(\phi_k))$$

where  $o_c$  is a syntactical operation on sentences associated with  $c$  (with  $O_c$  its propositional counterpart), and where  $\phi_1, \dots, \phi_k$  are well-formed formulas.

<sup>149</sup> Barwise & Etchemendy 1987 have a truth-theory that similarly quantifies over states of affairs and similarly fends off the slingshot objection owing to preservation of intensional structure. Moreover, theirs develops a theory for *type-free* truth predicates, thus bypassing the Tarskian requirement that we construct a hierarchy of languages in which conduct our semantical investigations. The difference between their approach and ours is that theirs only admits *atomic* states of affairs (with 1s and 0s as components), not complex states of affairs. In their view, like many others (see, for example, Pendlebury 1986), truth-makers are sets of atomic (obtaining) states of affairs, and complex, higher-level facts are determined by recursion on these atomic, basis-level facts.

slingshot arguments; we do not, as Barwise & Perry put it, “lose track” of subject matter that lead to slingshot type concerns.<sup>150</sup>

### 3.5.6 Addressing Three Popular Objections

The first objection, often raised, is that Tarski defined various predicates of the form ‘ $\lambda$  is true<sub>*L*</sub>’, and not a predicate of the form ‘ $\lambda$  is true-in-*L*’ for variable *L*. Furthermore, Tarski failed to furnish us with an ability to apply the concept to a new case (*cf.* Davidson 1996 (1999), p. 314). Hence, this objection continues, we cannot recognize that a Tarskian truth predicate is indeed a truth predicate *unless* we were to have some antecedent grasp of the general truth-concept – a grasp that Tarski necessarily failed to capture (*cf.* Dummett 1978, pp. xx-xxi). The best articulation of this objection is Hartry Field’s “Tarski’s Theory of Truth” (1972). There, Field maintains the criticism that the relation of satisfaction, being defined for particular languages, does not give us what is common to each definition and this is so because the basic clauses of the relation are enumeratively defined (or rest on the enumerative definitions of other semantic concepts – thereby, even more appositely for this objection, running counter to Tarski’s own expressed intentions). According to Field, Tarski eliminates the notion of truth by reducing it to other concepts, but does so without shedding light on the nature of truth. Field illustrates this by drawing on an analogy from chemistry. Consider what the notion of ‘*valence* of a chemical element’ would look like: “the integer that is associated with that element, which represents the sort of chemical combinations that the element will enter into.”<sup>151</sup> We could give a list-like definition of valence that listed every element and its associated valence, together with giving rules by which the valence of compound elements can be determined by their constituent, atomic elements. In this way, we could give an eliminative definition of valence in the form:

$$\forall E \forall n (E \text{ has valence } n \leftrightarrow E \text{ is potassium and } n \text{ is } +1, \text{ or } \dots, \text{ or } E \text{ is sulphur and } n \text{ is } -2) \quad ^{152}$$

Such a list-like definition does not demonstrate, however, what a definition of valence ought to give, *viz.* the physical properties of elements and how these properties make the elements bond in the proportions and manner that they do. Field’s claim is that Tarski gave us a definition of satisfaction that is similarly list-like and, thus, similarly unilluminating.

<sup>150</sup> Barwise & Perry 1981a. See also Restall 2004 for a similar argument.

<sup>151</sup> Field 1972, p. 362.

<sup>152</sup> *Ibid.*, p. 363.

This objection is answered by first pointing out that Tarski 1936a supplied an explication of the concept of truth (see 1983, pp. 152-154),<sup>153</sup> and explications are not normally achieved by provision of an explicit definition of the type ‘truth is...’. Rather, explications proceed step by step, from simple cases to more complex cases; the paradigmatic applications are extended to new (similar) cases which satisfy certain stipulated adequacy conditions<sup>154</sup> This type of research programme is familiar from explications of concepts like induction, confirmation and truthlikeness. As Niiniluoto remarks:

Tarski’s semantical approach to truth... follows this pattern of explication. His early paper defines truth for set theory and first-order logic, and later the programme is extended to cover new cases, like fragments of natural language, language of scientific theories, indexical expressions, vague statements, intensional languages etc.  
(Niiniluoto 1999a, p. 95)

Second, in the usual model-theoretic definition, Tarski gives a uniform definition that applies to *all* structures **M** that can interpret *L*’s descriptive vocabulary, and that truth with respect to **M** is entirely determined by the facts about **M** – which, of course, vary as we go from one structure **M** to any other **M**’.<sup>155</sup> Hence, no simple list will exhaust the content of a truth definition.<sup>156</sup>

The second objection is that Tarski’s theory fails to do justice to correspondentism owing to its employing the notion of satisfaction by *all (infinite) sequences of objects*, without discrimination. *All* true sentences are satisfied by *all* sequences of objects; if a sequence of objects satisfies one sentence when true, then they all do. In his later writings, Davidson goes as far as interpreting in this that “though this was not his intention, Tarski here indirectly vindicates Frege’s slingshot argument.”<sup>157</sup> Not being able to discern *which* sequence is doing the work is taken to significantly undermine Tarski’s success in discerning a “sufficiently precise and clear”<sup>158</sup> conception of the

---

<sup>153</sup> This is how many other authors interpret Tarski – see, for example Wolenski & Simons 1989a, Kirkham 1992, p. 170 and Niiniluoto 1999a.

<sup>154</sup> Cf. Carnap 1962.

<sup>155</sup> In the full blown calculus view, just like **M**, *L* too is universally quantified over so as to include the class of all first order languages (prefixing each basis clause with ‘If *L* contains an n-ary predicate R then...’, ‘If *L* contains constants *c*<sub>1</sub>, *c*<sub>2</sub>,... then...’, ‘If *L* contains an n-ary function symbol *f* then...’, etc).

<sup>156</sup> Moreover, as Carnap’s schema (C) (in the footnote above) from his semantical works in the 1930s demonstrates, there is a *general* schema underlying Tarski’s approach.

<sup>157</sup> Davidson 1999, p. 110.

<sup>158</sup> Cf. Tarski 1944, p. 54.

correspondence intuition expressed in the classical works Tarski aims to subserve. A.C. Grayling is an advocate of this objection. He says:

[Tarski's truth] definition seems to contain material for supporting a correspondence construal...it proceeds in terms of a definition of satisfaction, and satisfaction is a relation between sentences and sequences of objects. But the difficulty is that Tarski's definition states that true and false sentences are respectively satisfied by all sequences and none, with no appeal being made to *specific* sequences. (Grayling 1997, p. 156)<sup>159</sup>

This objection is unwarranted for the following reasons:

- (a) The fact that *all* true sentences are satisfied by *all* sequences of objects is an indication only that the truth of a sentence does not depend on the values of the free variables for the very simple reason that it does not have any. As the definition above shows, the only relevant members of the sequences are the relevant objects picked out by the variables when satisfied. To say that a sentence is true just in case it is satisfied by all infinite sequences is just a mathematical way of saying that the satisfaction for a closed formula does not depend on the valuation function chosen. The underlying logic has been misunderstood by Grayling *et al.*
- (b) The truth of atomic sentences need not be defined in terms of satisfaction; rather it is the intended domain that furnishes these sentences with the particulars and the properties/relations in which their truth consists.
- (c) Satisfaction can be defined *without* involving infinite sequences. Also, quantified sentences are (only) satisfied by the relevant members of the sequences of objects. For example, consider the existentially quantified sentence  $\exists x_2 A(x_2)$ , where  $A(x)$  has the intended interpretation ' $x$  is a Ph.D. student of Leśniewski'. According to Tarski, a sequence  $s$  satisfies  $\exists x_2 A(x_2)$  just in case some other sequence  $s'$ , agreeing with  $s$  in at most the 2<sup>nd</sup> place, satisfies  $A(x_2)$ , i.e. those that have Alfred Tarski – Leśniewski's only Ph.D. student – as their second member. The remainder of the sequence is irrelevant and made redundant.

---

<sup>159</sup> In a similar spirit, Haack (1978, p. 113) says: "*Tarski's definition of truth makes no appeal to specific sequences of objects, for true sentences are satisfied by all sequences and false sentences by none.*" Grayling's (and Haack's) can be viewed as being equivalent to the point that the model-theoretic definition, in and of itself, fails to provide explicit reference to facts, and is such that it cannot allow two distinct sentences of a language to correspond to two distinct facts when true. As pointed out above, section 3.5.5 addresses this type of criticism by presenting Tarski's theory in a fashion that does makes explicit reference to facts.

- (d) There is a sense in which the satisfaction for sentences *is* trivial. But the relation's *real* work is done by their operation on the open sentential functions that compose these sentences and, ultimately, on the atomic open sentential functions embedded within them. It is by making an appeal to the relations between sequences of objects and the atomic sentential functions that the satisfaction (and truth) of sentences is finally explicated. Also, in the atomic clauses of the recursive definition of satisfaction, different sequences of objects satisfy different atomic sentential functions and, thus, different sequences of objects *are* indeed distinguished.<sup>160</sup>

The third, and final, objection we deal with here is that Tarski's model-theoretic account applies *only* to uninterpreted, formalized languages whereas English is an interpreted, unformalized language. This sounds like a powerful objection but on examination it is no objection at all. While it is true, and universally acknowledged as such, that Tarski argued in favour of syntactically regimenting fragments of natural language so as to facilitate a truth definition – so that their syntax is to be made formally precise – and that it would not be possible to supply a truth definition for 'whole' natural languages owing to liar-type paradoxes resulting from semantic closure, we should note immediately that though working in formalized languages, Tarski's "real interest" was natural language:

I should like to emphasize that, when using the term 'formalized languages', I do not refer exclusively to linguistic systems that are formulated entirely in symbols, and I do not have in mind anything essentially opposed to natural languages. On the contrary, the only formalized languages that seem to be of *real interest* are those which are fragments of natural languages...or those which can at least be adequately translated into natural languages.  
(Tarski 1969, p.68. Emphasis added)<sup>161</sup>

We have already noted how Tarski explicitly restricted his attention to interpreted languages – in accordance with his heritage from the Lvov-Warsaw school: for Tarski (1936a), language is assumed to be interpreted in the domain of all objects (though the interpretation function is only implicitly present). The formalized languages Tarski employed are really more or less simple

---

<sup>160</sup> Despite his later reservations, Davidson understood this point in 1969: "*Truth is reached, in the semantic approach, by different routes for different sentences. All true sentences end up in the same place, but there are different stories about how they got there; a semantic theory of truth tells the story for a particular sentence by running through the steps of the recursive account of satisfaction appropriate to the sentence.*" (Davidson 1969 (1984), pp. 48-49).

<sup>161</sup> See also Tarski 1944 (1999), pp. 122-123, where Tarski remarks that his solution to the problem of truth is to be "rigorous" for "languages whose structure has been exactly specified" and "approximate" for natural languages. In dealing with natural languages, the portions we are interested in should be replaced by a syntactically well-specified language differing from the given natural language "*as little as possible*". This flatly contradicts Haack's claim that "Tarski...is thoroughly sceptical about the applicability of his theory to natural languages" (Haack 1978, p. 120).

models, but which have sufficiently many features shared with or approximating those of natural languages, and which can be used like experiments to test out certain claims and properties. The results have frequently yielded important information about the logical properties of *all* languages, formalized or not. This objection overlooks the vast amount of explanatory power you get as result, in particular about the nature of correspondence itself. Thus, Tarski's work shows that truth does, in quite a precise way, consist in correspondence with the facts; Gödel's incompleteness theorems tell us something we would never otherwise have known about the limits of deductive theorizing in general (in particular, no theory can capture all the truths about its domain of interpretation, and no theory can consistently prove its own consistency); Church's theorem tells us that we can't get a machine to decide mathematical truth; and Tarski's theorem not only implies Gödel's first incompleteness theorem as a corollary but tells us that truth is transcendental: it always outstrips the descriptive capacities of any language to characterize, whatever that language is.<sup>162</sup> English only appears to be a counterexample in being semantically closed, i.e. contains the wherewithal to talk of its own semantics, because the semantic closure entails inconsistency.<sup>163</sup>

The proposal here should be misunderstood as involving the claim that Tarski's is a good picture of semantics for *all possible* languages featuring true sentences; it would be too optimistic a proposal to claim that every possible language featuring true sentences is amenable to Tarskian semantics. What is an arbitrary sentence or language anyway? The notion of 'all possible languages' is virtually meaningless since there do not exist necessary and sufficient conditions which would be agreed on by all people for all times for any given thing to be a language. My – and I think Tarski's – aim is much more limited and I would claim scientific in spirit, which is to show how an authentic correspondence theory can be defended with respect to a model language that shares salient features with a natural language. As philosophers we should be *very* happy with a philosophically successful theory of truth for a (i) fragment of natural language, which is (ii) sufficiently idealized, and which (iii) allows only for purely extensional contexts (such as, for example, the language of some typical mathematical or scientific theory).<sup>164</sup> But, in any case, as picked out by Niiniluoto in the quotation

---

<sup>162</sup> Consider the language of Peano Arithmetic,  $\mathcal{L}_{PA}$ . Then, by Tarski's theorem, there is no formula  $A(x)$  in  $\mathcal{L}_{PA}$  such that in the standard model  $\mathbb{N}$ ,  $A(x)$  is satisfied by exactly the Gödel numbers true in  $\mathbb{N}$ .

<sup>163</sup> Kripke's work looks like another counterexample, but that is only because of a restriction imposed by semantic closure on its ability to express negation.

<sup>164</sup> Tarski's (early and late) work is extensional since it defines truth for sentences relative to the extensions of names and predicates. In Carnap 1963 (pp. 889-900), Carnap demonstrated how this can be reconciled with his own (1947) systematic theory of extensions and intensions (where notions like 'property' and 'proposition' are intensional entities expressed by predicates and sentences respectively): relative designation functions are defined in such a way that correlate linguistic expressions with their extensions in each 'possible world'. For example, for a given predicate  $P$  and possible world  $W$ , the designation function  $\text{Des}$  assigns a class of objects in  $W$ . These intensional entities – functions from possible worlds to extensions – became the key device in possible worlds semantics (cf. Hintikka 1975).



above from his 1999a paper, the philosophical research programme initiated by Tarski in the 1930s did lead to the development of semantics for more inclusive languages, such as for those concerning modalities and counterfactuals, i.e. possible worlds semantics (now universally considered as standard for such languages). Tarskian semantics was not abandoned for philosophical research programmes concerned with more inclusive languages.<sup>165</sup>

### 3.6 Closing Remarks

According to correspondentism, truth is a relational term meaning ‘correspondence’. This relation is comprehensively carved out by the semantical techniques of Tarski, and in a way that quietens Frege standing at bay against the reductionist programme. Tarski fine-grains the coarse-grained, folk theory of truth found in platitudes like ‘truth is correspondence with fact’ into a mathematically precise theory. Tarski ‘rehabilitates’ truth in the sense due to Popper: it is scientifically respectable to incorporate a notion expressing a significant language-world relation. The correspondence theory is reported by many to be vulnerable to the slingshot argument. This chapter establishes the validity of this argument, but also argues that all that the argument demonstrates is precisely the need for an appeal to the kind of conceptual apparatus provided by Tarski.

Truth has a nature, one expressing a (substantial) language-world relation. We have so far only considered substantialism about truth. Those who wish to advance an *insubstantialism* (or *deflationism*) about truth seek to deny the central presupposition – common to all substantialists – that truth has a nature. The following chapter outlines, and critically appraises, the theories of truth borne out of denying truth a nature.

---

<sup>165</sup> For successful applications not only to artificial formal languages but also to fragments of scientific, and of natural, languages, see, for example, Przelecki 1969, Tuomela 1973, Pearce 1987 and Niiniluoto 1987.

### 3.7 Appendix

#### (i) The Origin of the Correspondence Theory

Aristotle *Metaphysics*,  $\Gamma$ , 7, 1011<sup>b</sup>25 is often cited as the original articulation of the correspondence thesis.<sup>166</sup> Marian David (David 1994, 2005) disagrees, suggesting instead that we ought to locate the first proper citing in Thomas Aquinas's *Questiones Disputatae De Veritate*, Q.1, A.1&3:

Veritas intellectus est adaequatio intellectus et rei; secundum quod intellectus 'dicit esse quod est vel non esse quod non est'.

("The truth of the intellect is the agreement between intellect and thing; insofar as the intellect 'says of that which is that it is or of that which is not that it is not'.")

However, Wolfgang Künne writes that we could legitimately date the true origins of the correspondence interpretation to the second century BC – to Carneades.<sup>167</sup> As head of the Academy, Carneades is reported to have taught that "a presentation is true when it is *in accord with* the object presented, but false when it is in discord with it".<sup>168</sup>

#### (ii) The (Tarskian, 1936) Route from Satisfaction to Truth

Consider an object language  $L$  with the following characteristics:

Vocabulary of $L$ :	<i>parentheses</i>	$)$ , $($
	<i>variables</i>	$v_1, v_2, \dots, v_k, \dots$
	<i>logical connectives</i>	$\wedge$ (and), $\neg$ (not)
	<i>quantifier</i>	$\forall$ (for all)
	<i>individual constants</i>	$c_1, c_2, \dots, c_k, \dots$
	<i>function symbols</i>	$f^n_k$ ( $n$ and $k$ are any positive integer)

---

<sup>166</sup> Some feel, however, that despite invoking *some* relation to reality this definition – to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true – does not do full credit to the spirit of correspondence: the relation is not made explicit and, moreover, there is no specification of what on the part of reality is responsible for establishing what truth consists in. This concern with Aristotle's strict allegiance to correspondentism is misplaced. Elsewhere, in *Categories*, Aristotle does give a more faithful version: "*The fact of the being of a man carries with it the truth of the proposition that he is...for if a man is, the proposition wherein we allege that he is, is true...the fact of the man's being does seem somehow to be the cause of the truth of the proposition, for the truth or falsity of the proposition depend on the fact of the man's being or not being.*" (Aristotle *Categories*, 14<sup>b</sup>14-21).

<sup>167</sup> Künne 2003.

<sup>168</sup> Emphasis added. Quote taken from Künne 2003, p. 103, who in turn references it to Plotinus c.250.

predicate symbols  $P^n_k$  ( $n$  and  $k$  are any positive integer)<sup>169</sup>  
 one binary relation symbol  $\equiv$  (identity)

Inductive Definition of *Well-Formed Term* of  $L$ :

- (i) A variable is a well-formed term.
- (ii) A constant symbol is a well-formed term.
- (iii) If  $f^n_k$  is a function symbol and  $t_1, \dots, t_m$  are terms then  $f^n_k(t_1, \dots, t_m)$  is a well-formed term.
- (iv) Only expressions obtained by a finite number of applications of (i)-(iii) are well-formed terms.

Inductive Definition of *Well-Formed Formula* of  $L$ :

- (i) If  $t_1$  and  $t_2$  are well-formed terms then  $t_1 \equiv t_2$  is a well-formed (atomic) formula.
- (ii) If  $P^n_k$  is a predicate symbol and  $t_1, \dots, t_m$  are terms then  $P^n_k(t_1, \dots, t_m)$  is a well-formed (atomic) formula.
- (iii) If  $\phi$  is a well-formed formula then so is  $\neg\phi$ .
- (iv) If  $\phi$  and  $\psi$  are well-formed formulas then so is  $\phi \wedge \psi$ .
- (v) If  $v$  is an arbitrary (free) variable and  $\phi$  is a well-formed formula then  $\forall v\phi$  is a well-formed formula (the variable  $v$  is *bound* by the quantifier).
- (vi) Only expressions obtained by a finite number of applications of (i)-(v) are well-formed formulas.

Definition of a *Sentence* of  $L$ :  $\Sigma$  is a sentence of  $L$  if and only if  $\Sigma$  is a well-formed formula of  $L$  with no free variables.

Let  $L$  be given an interpretation  $\mathbf{M}$  with domain  $D$ . In this model, we fix some (ordered) denumerable sequence of objects or individuals  $s = \langle s_0, s_1, s_2, \dots \rangle$  that will get assigned to the (free) variables  $v_0, v_1, v_2, \dots$ , respectively. Let  $\Phi$  be the set of all denumerable sequences of elements of  $D$ . To help us define what it means for a sequence  $s = \langle s_0, s_1, s_2, \dots \rangle$  in  $\Phi$  to satisfy a well-formed formula in  $\mathbf{M}$ , we need to define a function  $s^*$  that assigns to each term  $t$  an element  $s^*(t)$  in  $D$ :

<sup>169</sup> The positive integer  $n$  indicates the number of arguments, whereas the subscript  $k$  is an indexing number distinguishing different predicate and function symbols with the same number of arguments.

- (i) If  $t$  is a variable  $v_j$ , let  $s^*(t)$  be  $s_j$ .
- (ii) If  $t$  is an individual constant  $c_j$ , then  $s^*(t)$  is the interpretation  $(c_j)^{\mathbf{M}}$  of this constant.
- (iii) If  $f_k^n$  is a function symbol,  $(f_k^n)^{\mathbf{M}}$  is the corresponding operation in  $D$ , and  $t_1, \dots, t_m$  are terms, then  $s^*(f_k^n(t_1, \dots, t_m)) = (f_k^n)^{\mathbf{M}}(s^*(t_1), \dots, s^*(t_m))$ .<sup>170</sup>

Now we can proceed to a recursive definition of satisfaction by  $s$ :

- (iv) If  $\phi$  is the well-formed, atomic formula  $P_k^n(t_1, \dots, t_m)$  and  $(P_k^n)^{\mathbf{M}}$  is the corresponding  $n$ -place relation of the interpretation, then  $\phi$  is satisfied by sequence  $s$  if and only if  $(P_k^n)^{\mathbf{M}}(s^*(t_1), \dots, s^*(t_m))$ .<sup>171</sup>
- (i)  $\neg\phi$  is satisfied by  $s$  if and only if  $\phi$  is not satisfied by  $s$ .
- (ii)  $\phi \wedge \psi$  is satisfied by  $s$  if and only if  $\phi$  is satisfied by  $s$  and  $\psi$  is satisfied by  $s$ .
- (iii)  $\forall v_i \phi$  is satisfied by  $s$  if and only if each sequence  $s'$  differing from  $s$  in at most the  $i$ -th component satisfies  $\phi$ .

What we need now is a lemma effectively saying that the value of any term  $t$  at  $s_0, s_1, s_2, \dots$  and the satisfaction (or otherwise) of any formula  $\phi$  by  $s_0, s_1, s_2, \dots$  depends *only* on the those values of  $s_i$  for which  $v_i$  is a free variable, i.e. it is *independent* of all the other members, and the length, of the sequence.<sup>172</sup> From this, we obtain the following corollary:

Let  $\phi$  be any formula with no free variables, and whose bound variables are among  $v_0, \dots, v_p$ .  $\mathbf{M}$  satisfies  $\phi$ , i.e.  $\mathbf{M} \models \phi$ , just in case  $\phi$  is satisfied in  $\mathbf{M}$  by some (or, equivalently, every) sequence  $s_0, \dots, s_p$ .

Whence we get our definition of truth:

$$\forall x [x \text{ is true-in-} L \text{ iff } (x \text{ is a sentence-in-} L \text{ and } x \text{ is satisfied-in-} L \text{ by all sequences } s)]$$

<sup>170</sup> Intuitively,  $s^*(t)$  is the element of  $D$  obtained by substituting, for each  $j$ , a name  $s_j$  for all occurrences of  $v_j$  in  $t$  and then performing the operations of the interpretation corresponding to the function letters of  $t$ .

<sup>171</sup> I.e., if the  $n$ -tuple  $\langle s^*(t_1), \dots, s^*(t_m) \rangle$  is in the relation  $(P_k^n)^{\mathbf{M}}$ . If, for example, the domain of interpretation is the set of real numbers, the interpretation of  $P^2_I$  is the relation  $\leq$ , and the interpretation of  $f^1_I$  is the function  $e^v$ , then  $P^2_I(f^1_I(v_2), v_5)$  is satisfied by sequence  $s = \langle s_0, s_1, s_2, \dots \rangle$  if and only if  $e^{s_2} \leq s_5$ .

<sup>172</sup> The proof is given by induction on the complexity of formulas. See Proposition 1.3.16 of Chang & Keisler 1973.

Tarski worked in the calculus of classes, where ' $\subseteq$ ' is the only predicate symbol. His recursive definition of satisfaction is given in 1936a by definition 23:

**DEFINITION 22:** The sequence  $f$  satisfies the sentential function  $x$  if and only if  $f$  is an infinite sequence of classes and  $x$  is a sentential function and  $f$  and  $x$  are such that either:

- ( $\alpha$ ) there exist natural numbers  $k$  and  $l$  such that  $x = 't_k \subseteq t_l'$  and  $f_k \subseteq f_l$ ;
  - ( $\beta$ ) there is a sentential function  $y$  such that  $x = '\neg y'$  and  $f$  does not satisfy the function  $y$ ;
  - ( $\gamma$ ) there are sentential functions  $y$  and  $z$  such that  $x = 'y \vee z'$  and  $f$  either satisfies  $y$  or satisfies  $z$ ;
  - ( $\delta$ ) there is a natural number  $k$  and sentential function  $y$  such that  $x = '\forall_k y'$  and every infinite sequence of classes which differs from  $f$  in at most the  $k$ -th place satisfies the function  $y$ .
- (Tarski 1936a (1983), p. 193. I have updated the notation slightly.)

Whether a formula is satisfied by a given sequence depends on the correspondence between the free variables of the formula and the terms of the sequence, co-ordinated through indices. From a proof of a lemma showing that whether a sequence  $s$  satisfies a formula  $F$  depends only on the values of  $s$  on the free variables of  $F$ , it immediately follows that for closed formulas either every infinite sequence satisfies it, or none do. And this is Tarski's definition 23.<sup>173</sup>

### (iii) Correspondence as Homomorphism<sup>174</sup>

Tarski truth-definitions can be expressed in the following form:

Consider a first-order language  $L$  containing constants  $a, b, c, \dots$  and  $n$ -place relation symbols  $P, Q, R, \dots$ . Let  $I$  be the fixed interpretation for  $L$  and let  $C$  be the set of constants (assumed enough to name all the elements of the domain  $D$  to simplify this little example) and the relation symbols of  $L$ . Let  $F$  be the set of formulas of  $L$ , defined as the closure of the atomic formulas under the application of the operators  $\neg$  (unary),  $\wedge$  (binary), and the infinitely many operators  $\wedge x$  (unary) for each variable  $x$  (universal quantifiers). These are assumptions always made in standard logical semantics, as any logic text will confirm, and are merely part of the modelling strategy. As with many model-simplifying moves, these conditions can be relaxed at the cost of somewhat greater complexity.

<sup>173</sup> " $x$  is a true sentence – in symbols  $x \in Tr$  – if and only if  $x \in S$  [the set of meaningful sentences] and every infinite sequence of classes satisfies  $x$ ." (Tarski 1936a (1983), p. 195). As pointed out above, the use of infinite sequences is only one (and not the most natural) way of defining satisfaction: you can instead use valuation functions with finite domains that assign values to the variables contained in the formulas. Moreover, if you assume that your language has names for all elements of the domain you can define truth recursively on sentences, i.e. closed formulas, themselves. See the introductory chapter of Chang & Keisler 1973.

<sup>174</sup> See Dunn & Hardegree 2001, p. 15.

Let  $D^n$  be the  $n$ th Cartesian product of  $D$  with itself. It can be represented as an ordered pair  $(D, v)$ , where  $v$  is a function from  $C \cup F$  into  $D \cup D^2 \cup \dots \cup D^n \cup \dots \cup \{0,1\}$ . Let  $Id$  be the 2-place function from  $D^2$  into  $\{0,1\}$  such that  $Id(u,v) = 1$  if  $u = v$  and  $Id(u,v) = 0$  if not (i.e.  $Id$  represents the identity relation on  $D$ ). Let  $R$  be a function from  $D^n$  into  $\{0,1\}$ , let  $\sim$  be the function from  $\{0,1\}$  into  $\{0,1\}$  such that  $v(0) = 1$  and  $v(1) = 0$ . Define  $v$  inductively as follows:

1.  $v(c) \in D$  for every constant  $c$  of  $L$
2.  $v(c=d) = Id(v(c), v(d))$
3.  $v(R(a_1, \dots, a_n)) = v(R)(v(a_1), \dots, v(a_n))$
4.  $v(\neg(A)) = \sim(v(A))$ , where  $A$  is a well-formed formula
5.  $v(\wedge(A,B)) = \wedge(v(A), v(B))$  where  $\wedge$  denotes the greatest lower bound and  $A, B$  are well-formed formulas
6.  $v(\wedge x(A)) = \wedge(v(A(a/x)))$  for every  $a$  in  $L$

Note that  $v$  is a union of homomorphisms, and a sentence of  $L$  is true in  $I$  just in case its value is 1 under  $v$ . In fact, we can represent each of the homomorphisms by a commutative diagram: e.g. the one for  $\neg(A)$  is:

$$\begin{array}{ccc}
 & \neg & \\
 F & \xrightarrow{\quad} & F \\
 v \downarrow & & \downarrow v \\
 \{0,1\} & \xrightarrow{\quad} & \{0,1\} \\
 & \sim &
 \end{array}$$

# CHAPTER 4

## DEFLATIONISM

It cannot be that axioms established by argumentation provide for the discovery of new works, since the subtlety of nature is greater more times over than the subtlety of argumentation.

**Francis Bacon, *Novum Organon***

The umbrella term *Deflationism* is a doctrine understood in a number of different ways by different philosophers. No attempt at specifying what might be the central deflationary characteristics would invite anything like universal agreement; no *one* deflationist would commit herself to any proposed (exhaustive) list of commitments.<sup>175</sup> One might think that picking out what each deflationary theory of truth has in common would furnish us with the core principles of a coherent doctrine of deflationism. However, there are too many alternative, mutually exclusive, deflationary conceptions of truth for this to present itself as a credible strategy: the result would approximate the empty set. Deflationism is best characterised, if you like, as more of a philosophical tendency. Fundamentally, those succumbing to this tendency do so owing to the belief that the concept of truth is “a relic of a bygone age”, as Bertrand Russell famously (and incorrectly) once said of the notion of cause in modern day physics.<sup>176</sup> Deflationism about truth is thus not a single creed whose distillation commands universal consent. Rather, it is a *pot-pourri* of (strongly non-equivalent) ideas. Broadly, the spectrum is this:

- Truth is not a genuine property; as with the notion of ‘existence’, the surface grammar of language misleads us into thinking that truth is a property ascribed by a predicate
- The truth predicate is redundant; if anything, to call something true is an illocutionary act of endorsement
- The statements ‘*S* is true’ and ‘*S*’ are (logically, or cognitively, or necessarily) equivalent
- Truth may be a real property, but it is indefinable in that what underwrites our understanding of the property is our disposition to accept each *T*-schema instance; if you prefer, truth is *implicitly defined* by that schema

---

<sup>175</sup> Maybe parts of any such proposed such list. But the whole point is that the parts subscribed to would vary among deflationists.

<sup>176</sup> Russell 1912b.

- The concept of truth is a logico-mathematical device of generalization or semantic ascent, allowing us to express infinite (logical) conjunctions and disjunctions
- The truth predicate is a linguistic device: of ‘disquotation’, of sentence denominalization, or an anaphora-inducing, prosentential operator; it also allows us to restore the structure of a complete sentence
- The theory of truth is neutral with regard to epistemological/metaphysical questions and the concept of truth (on its own) has no theoretical explanatory power
- As the truth property, if any, does not exist independently of (and above) the logical and practical functions performed by the truth predicate, we need to minimize the employment of philosophically-laden conceptual resources in explaining it

This chapter attempts to elicit what the core of deflationism might be. It begins by summarizing – as far as is possible – what the constituents of this core are. Then, since deflationists can be viewed as grouped into either one of two camps – those that see the *T*-scheme as fundamental, and those that do not – the latter of these groups is dealt with first. For *T*-scheme based deflationists, one might propose the following three deflationary positions as currently dominating the literature: redundantism, disquotationalism and minimalism, and each is then subsequently treated separately. This chapter ultimately finds the deflationary attitude to truth untenable.

There are numerous references for deflationism, together with authors’ attempts at identifying what this philosophy consists in by examining the principal actors involved. They include Halbach 2004, Armour-Garb, B. & Beall, J.C., 2005a, Candlish, S. & Damnjanovic, N. 2007, Stoltjar, D. & Damnjanovic, N. 2007, Kirkham 1992 and Künne 2003. This chapter has derived inspiration from many of these.

#### **4.1 *Introductory Remarks***

Deflationists oppose the idea that the question ‘What is (the nature of) truth?’ is in any sense meaningful for the simple reason that they (collectively) reject the claim that there is indeed a nature to truth; there is thus no philosophical value added to posing such questions. If there is no such nature, then the correspondentist is erroneously attempting to unearth or uncover its features. Any attempt at reductively analyzing the notion of truth in the following manner:

$x$  is true if, and only if  $x$  is  $F$



is fraught with difficulties, and unsurprisingly so for the deflationist, since no such reductive analysis identifying truth's constituents is at all possible. This is not to say that the deflationist does not take away from the correspondentist any valuable lessons. Many card-carrying deflationists accept that the correspondence intuition is a significant constraint on an adequate theory of truth – they adopt exactly the same methodological attitude in attempting to do justice to our pre-analytical understanding.<sup>177</sup> Where the deflationist parts company with the correspondentist is her view that the correspondence account – which for her naïvely holds onto to pseudo-explanatory, weighty (or substantial) concepts like 'correspondence' and 'fact' – should be trimmed, or deflated.<sup>178</sup> In so doing, we would leave bare truth's insubstantiality.

Deflationism characterizes itself as a negative thesis in contrast with the substantialist presuppositions of correspondentism, which it aims to deflate. Some have argued, for instance Paul Horwich and Michael Williams, that part of this contrast involves the claim that deflationism employs a *minimum* of conceptual resources.<sup>179</sup> But it is actually difficult to see that herein lies a defining contrast. The argument is that, while the correspondentist appeals to deeply philosophically-laden conceptual resources such as 'states of affairs obtaining', 'corresponding to facts', 'representation' and their ilk, the deflationist means only to get by with the fewest possible theoretical constructions. This attitude is most telling in the following deflationist's remark:

[D]eflationists...are united in their intention to explain the workings of truth-talk in a theoretically economical way...This is to be expected. If the function of truth-talk is merely expressive, it is reasonable to suppose that the mechanism of truth can be explained using limited theoretical resources.  
(Williams 2002, p. 148)

If part of being a deflationist about truth involves invoking only a minimum of explanatory resources, then there is no reason why the correspondentist could not either.<sup>180</sup> After all, both share *exactly* the same methodological attitude in attempting to fine-grain our already coarse-grained, pre-theoretical understanding of truth and appeal to the same schematic representation of what the truth

---

<sup>177</sup> Cf. Douven & Hindriks 2005 who argue that the deflationist need not be so constrained. If not, then it is difficult to see in what sense the deflationist is a theorist about *truth*. Both correspondentists and deflationists attempt to fine-grain our pre-theoretic, folk understanding of the notion. And this understanding is fundamentally underwritten by the correspondence intuition. See Christopher Hill's "The Marriage of Heaven and Hell: Reconciling Deflationary Semantics with Correspondence Intuitions" (2001) which also discusses this.

<sup>178</sup> This contrast between deflationism and correspondentism was discussed in 3.2.1.

<sup>179</sup> See Horwich 1998a, p 6. David 1994 p. 4 also points out that according to the deflationist "[t]he correct explanation of truth...requires less extravagant resources" than those found in the traditional substantive accounts.

<sup>180</sup> Mou 2000, pp. 266-268 argues that this is an important metaphilosophical point bearing on the deflationist-correspondentist distinction.

of a claim consists in.<sup>181</sup> This might explain why, in some accounts of truth, deflationism is articulated *not* in opposition to substantialist accounts of truth *tout court* – for this would include epistemic theories – but in opposition only to non-epistemic accounts that conform to our pre-theoretic grasp of truth. Both share the presupposition that philosophy cannot be *revisionary* in the sense of countering this kind of understanding; this methodology counsels us not to scratch where it does not itch. Blackburn calls this approach to philosophy ‘quietism’.<sup>182</sup> It is clear that this owes much to the later philosophy of Wittgenstein, who claimed that, instead of solution, what philosophical problems really require is therapy.<sup>183</sup> There are radical versions of quietism, versions according to which philosophy has not much to say at all about its traditional concerns. In the context of the present debate however, it is reasonable to suppose that both parties involved – substantialists and insubstantialists alike – are *moderate* quietists, in that we ought not to revise our ordinary understanding of truth *unless* there are ineluctable theoretical reasons compelling us to do so.<sup>184</sup> Appealing to minimal explanatory resources is a way open both to deflationists and to correspondentists. If there is a contrast, it does not lie in the number of conceptual resources called upon by both theses.

From the *negative* thesis that we should do without the obscure, mysterious and (worse) question-begging constructions of correspondentism, the deflationist takes a *positive* turn in her outlook. The positive thesis is that no more information about truth can be garnered than that already gotten by adopting the schemata:

the sentence ‘*p*’ is true if and only if *p*, or  
the proposition that *p* is true if and only if *p*

These adoptions are such that the *T*-scheme is to be considered *conceptually basic*. Davidson (though not a deflationist) said in this connection:

The reason Convention **T** is acceptable as a criterion of theories [of truth] is that...*T*-sentences are clearly true (pre-analytically) – something we could recognize only if we already partly understood the predicate

---

<sup>181</sup> Of course, how to understand the *T*-schema is differently accounted for by deflationists and correspondentists.

<sup>182</sup> Blackburn 1984, p. 146.

<sup>183</sup> Although see Blackburn 2005, pp. 129-136, on why heralding Wittgenstein as the patron saint of quietism would be a “*bizarre misreading*” of his later works. See also O’Leary-Hawthorne & Price 1996 for a discussion of the quietist approach.

<sup>184</sup> And even then, it is doubtful whether moderate quietism ought to be the *global* attitude to take with respect to philosophical problems.

'is true' ...  
(Davidson 1977 (1984), p. 218)

It is perhaps here that we ought to pick up on a suggestion made by Bo Mou (Mou 2000). At the metaphilosophical level, both deflationists and correspondentists adopt the quietist, non-revisionary attitude. However, the deflationist locates the primary pre-theoretical understanding of truth in the logico-linguistic functions it performs. This is a *necessary* part of what it means to be a deflationist about truth. The same clearly does not apply to the correspondentist; she gives primary status to the fact that true sentences describes things as they really are, or, in other words, tell it like is.

The primacy of the *T*-scheme accounts for why most deflationists ascribe a modal status to its instances – that they be considered *necessary* or *a priori*. Hence, the only intelligible question to ask with respect to the notion of truth is 'What (conceptual) role does the truth predicate play in our language?' and, clearly, this role, given the modal status of the *T*-sentences, will be underwritten by the *T*-scheme. What, then, is this role that an observation of linguistic behaviour tells us is played by the concept of truth? The deflationist highlights the following:

- (a) Blind Ascriptions and Endorsements: Not knowing what Michael Dummett has just said, and supposing he just said 'We ought to reason intuitionistically', then, given the relevant instance of the *T*-scheme, I can indirectly make the same assertion by saying 'What Michael Dummett said is true'. Relatedly, I am able to endorse or commend Michael Dummett's last claim by employing the truth predicate. Note, that the correspondentist does not deny any of these truth-theoretic roles. Where the correspondentist differs is in her explanation of how truth is conferred these abilities. She would typically argue that these abilities are conferred owing to some fact about the *nature* of truth. As deflationists deny any presupposition to the effect that truth has a nature, she would instead say that it is in virtue of the *T*-scheme and nothing more than that, that enables truth to play these roles. She might even insist that, even if truth admits of reductive analysis in the way rejected by the deflationist, we need not look further than the above schema to explain these linguistic roles.
- (b) Generalizations: Evidently, it is possible to make indirect assertions *without* the concept of truth. But the point that deflationists wish to stress is that truth facilitates this function in a finitary (and, therefore, accessible) manner. For example, I could indirectly assert (or endorse) what Michael Dummett said by using the following infinite conjunction:

(If what Michael Dummett said is 'snow is white' then snow is white) &

(If what Michael Dummett said is ‘grass is green’ then grass is green) &  
(If what Michael Dummett said is ‘Bandersnatches are frumious’ then Bandersnatches  
are frumious) &...

or, the following infinite disjunction:

(What Michael Dummett said is ‘snow is white’ and snow is white) or  
(What Michael Dummett said is ‘grass is green’ and grass is green) or  
(What Michael Dummett said is ‘Bandersnatches are frumious’ and Bandersnatches are  
frumious) or...

or, by simply employing the truth predicate as above. The advantage, of course, with the latter (the point pushed by deflationists) is that we can *finitarily* capture the intending meaning. In other words, the infinite disjunction/conjunction is of the same logical strength; they are equivalent in this sense. The situation is exactly the same when dealing with quantificational phrases. For example, we could assert the infinite conjunction of instances of the Law of Excluded Middle by asserting the single sentence “Every instance of ‘p or not-p’ is true”. The truth predicate affords language-users the ability to finitarily *express* a potentially infinite conjunction/disjunction, i.e. furnish us with a sentence that has the same logical power. And what underwrites truth the ability to perform these functions is the equivalence between a sentence ‘p’ and the sentence, “‘p’ is true”, i.e. the *T*-scheme.

There is nothing, says the deflationist, of any philosophical interest pertaining to truth in addition to identifying the logico-linguistic functions it performs. No singular, unifying explanation to which truth is reducible is possible. Deflationists, therefore, reject this ‘substantialist’ metaphilosophical presupposition. Truth cannot exist independently of these functions and, indeed, these functions are determined, or licensed, by the *T*-scheme, which is to be considered conceptually basic. Owing fundamentally to the primacy of the *T*-scheme, the abilities truth possesses cannot tell us anything more about the non-semantical world than what we have already gained without it. A manifestation of this is the equipollence in logical strength of the infinite conjunctions/disjunctions and their finitary truth-theoretic counterparts.

The observation that the *T*-scheme is to be viewed as conceptually basic permits us to immediately discard any deflationary theory of truth – or *any* theory of truth, for that matter – failing to endorse this commitment. This is owing to the fact that a fundamental constraint of adequacy on any

candidate theory of truth is that it do justice to our pre-theoretical commitment to the idea that a proposition is true if and only if it corresponds with fact. Strawson's illocutionary theory, C.J.F. Williams' redundancy theory and the Prosential theory all fail to meet this constraint, but nonetheless, perplexingly, are among the theories that dominate the literature on deflationism. We ought to first cursorily examine these theories in turn, partly due to the punctilious purpose of paying homage to their presence in the literature, and partly also because their critical examination exposes some important philosophical lessons. We then turn our attention to deflationary theories that view the *T*-scheme as fundamental.

## 4.2 Deflationary Theories Not Committed to the *T*-scheme

### 4.2.1 Strawson's Illocutionary Theory

On Strawson's view, what we do when uttering a truth ascription is perform a certain kind of speech-act: that of endorsing or signalling our agreement with something.<sup>185</sup> Analogously with emotivism in ethics – where to make utterances of the form 'torture is wrong' only indicates the speaker's attitude to torture, and (significantly) does not, despite appearances, predicate 'is wrong' of torture – Strawson's view is that to utter sentences like "'Snow is white' is true" or 'What the policeman says is true' is not, strictly speaking, to say anything at all; it is, rather, *doing* something, namely confirming, very much like uttering 'confirm that snow is white' or 'Hooray to whatever the policeman says', which are more like gestures. Hence, to utter "'Snow is white' is true" is not to say anything about the sentence 'Snow is white'. Indeed, as is widely acknowledged, Strawson makes the stronger claim that the truth predicate is not applicable to anything whatsoever, from which it follows that it cannot be applied to sentences:

[T]he phrase 'is true' is not *applied to* sentences; for it is not *applied to* anything.  
(Strawson 1949, p. 84)

Truth ascriptions are, therefore, *nonassertoric performative* utterances:

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 said; but (except where we are implicitly making a meta-statement [*viz.*, that so-and-so made a statement], in making

<sup>185</sup> The titles *Performative Theory of Truth* and *Expressivist Theory of Truth* are sometimes also given to this deflationary position.

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...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. (Strawson 1949, p. 92)

Moreover, given the illocutionary role played by truth, attributions of truth, on this view, typically take place in a context – one where someone first utters a statement that is then endorsed or confirmed by someone else. The context is important in allowing one to apply truth, to gesture their agreement to what is said. Both Ramsey (at one time) and Carnap (at another) held similar views to Strawson's here. In his Critical Notice of Wittgenstein's *Tractatus Logico-Philosophicus*, Ramsey points to the emotive or stylistic uses to which 'true' is put. In his *Introduction to Semantics*, Rudolf Carnap, whilst endorsing the semantic equivalence of the claims '*p*' and '*p* is true', also says that they have:<sup>186</sup>

...different features and different conditions of application; from this point of view we may e.g. point to the difference between these two statements in emphasis and emotional function. (Carnap 1942, p. 26)

There are some serious objections and drawbacks to this view on truth and the following three are, in my view, the most telling:

- (a) There is simply no argument offered for this deflationary thesis. Rather, as Richard Kirkham (1992 p. 311) points out, Strawson presupposes the following *Ockham's Razor* principle: one ought only postulate entities when not doing so disables us from adequately explaining the semantics and syntax of language use. Since, on this view, linguistic analysis shows that the use of the truth predicate fails to name a property, to argue that there can therefore be no such property is a highly dubious inferential leap. It is no less reasonable to infer that the linguistic analysis *conceals* the identity of the truth-property and truth-bearer than to infer that it *reveals* that there is no such property or entities corresponding to its bearer.<sup>187</sup>

---

<sup>186</sup> Ramsey 1923, Carnap 1942. Frege makes this point too: "The truth claim arises...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...the same thought as the simple '5 is a prime number'." Frege 1892 (1980), p. 64.

<sup>187</sup> On this point, Kirkham says, "This kind of assumption about how to discover metaphysical truths has a venerable history in philosophy, but most of those who have relied on it implicitly would cough nervously as seeing it laid out baldly like this." (Kirkham 1992, p. 71)

- (b) If truth-ascriptions are illocutionary acts only, then it is difficult to see how they could participate in arguments we standardly accept as valid. If the truth-ascription ‘Oscar’s claim is true’ is an *act*, then it cannot play the role of a *premise* in the following argument:

Oscar’s claim is that snow is white; Oscar’s claim is true; *ergo*, snow is white.

As this argument is indubitably valid, it would seem that Strawson’s performative theory of truth must be wrong. Truth predications are, indeed, associated with speech-acts, like that of conceding, but, as Warnock 1964 observes, we do so by attributing the property of truth to the claim in question, not instead of it.<sup>188</sup>

- (c) In 1950, Strawson, whilst still maintaining the performative theory, did concede that certain non-linguistic conditions must obtain to equip truth-ascriptions the ability to perform their function as acts of endorsement. One can’t say, for example, “‘Penguins waddle’ is true” unless it is a fact that penguins waddle: “Certainly, we use the word ‘true’ when the semantic conditions...are fulfilled; but we do not, in using the word, *state* that they are fulfilled”.<sup>189</sup> But if the semantic conditions are not being stated, then how are they relevant to the use of the truth predicate, as understood by the illocutionary theory? Surely, in making the (honest) gesture of signalling one’s agreement, one is *saying* something about the relevant facts of the matter. Strawson 1964 actually concedes this point – that, whilst still performing an action, one is also saying something in making utterances like ‘*p* is true’.

#### 4.2.2 C. J. F. Williams’ Redundancy theory

Williams’ principal reference here is his 1976 *What is Truth?*. All commentaries on Williams refer to Williams’ views by referring to this particular source. In *What is Truth?* Williams says:

The doctrines for which I argue are, to a large extent, inherited from the discoveries made in Philosophical Logic by the late Arthur Prior.  
(Williams, C.J.F. 1976, p. *xiii*)

Later in his book, Williams confirms that:

...the type of analysis of truth we are offering is due in great measure to the methods employed in this

---

<sup>188</sup> Horwich 1998 and Geach 1960 make these objections too.

<sup>189</sup> Strawson 1950, p. 44.

area by the late Arthur Prior.  
(Williams, C.J.F. 1976, p. 3)

It is therefore prudent to begin our exposition of Williams' theory with those aspects of Prior's philosophy Williams viewed as so influential. The reference here is Prior's 1971 *Objects of Thought*. As the title suggests, Prior sought to understand both 'what we think' and 'what we think of'. In the course of doing so, he undermines (according to Williams) the correspondence theory of truth by questioning both the notions of 'proposition' and of 'fact'. According to Prior, if we begin with the idea that the primary truth-bearers are whatever the objects of thought are, then it would be mistaken of us to think that our believing something is a relation involving sentences. Instead, believing must be a relation involving propositions. But there are no such things as propositions, for propositions are *logical constructions*, i.e. sentences which have grammatical subjects appearing to designate propositions mean no more than equivalent sentences in which no expression plays this role. Thus, 'The proposition that Madrid is in Kenya is false' means no more than 'It is not the case that Madrid is in Kenya':

'Propositions are logical constructions' was first said as a summing-up of this theory [viz. Russell's multiple relation theory of judgment]. It meant that statements which appear to be about people and propositions are really about people and quite other things, so that it is not necessary to suppose that there really are such things as propositions.  
(Prior 1971, p. 8)

Hence, 'The proposition that Madrid is in Kenya is believed by Joshua' means that the property of being in Kenya as far as Joshua is concerned belongs to Madrid; it does not, according to Prior, ascribe the property of being believed by Joshua to some proposition. Facts, similarly, are logical constructions:

...facts and true propositions alike are mere 'logical constructions'...and...they are the *same* 'logical constructions' (to have 'true propositions' and 'facts' is to have *too many* logical constructions).  
(Prior 1971, p. 5)

If all possible objects of thought are dismissed, then truth cannot be a genuine predicate at all and there are no objects that we can appropriately say bear truth. Williams was deeply influenced by all this, favouring blind ascriptions of truth to explicit ascriptions as his arena for analysis. This did not mean that Williams denied the significance of explicit, or transparent, ascriptions of truth. He only wished to stress that, owing to their being the most common and fundamental kind, blind ascriptions of truth are those we ought to analyse and then, presumably, that analysis would carry



over to ascriptions of truth where the sentence concerned is explicitly displayed.<sup>190</sup> Blind ascriptions, such as ‘What Dummett says is true’, would then be explained thus (cf. Williams 1976, p. 38):

$$\exists p \forall q ((p = q \leftrightarrow \text{Dummett says that } q) \wedge p)$$

Now, according to Williams, despite appearances to the contrary the nominal phrase ‘What Dummett says’ does not name anything; it is to be understood rather as behaving just like a quantifier. The grammatical subject of the sentence – ‘What Dummett says’ – fails to name anything, anymore than ‘What the postman brought’ does in ‘What the postman brought is on the mantelpiece’; ‘What the postman brought’ and ‘What Dummett says’ are (in the language of Russell and Whitehead’s *Principia*) “incomplete symbols”, much like the quantifiers. The phrase ‘is true’ allows us to complete the structure of a sentence without giving it the content of a completed sentence. Hence, the facts of English create the illusion of making us think ‘What Dummett says is true’ concatenates a name with a predicate, since it contains neither.

We can note immediately that the criticism levelled at Strawson above {criticism (a)} applies here too. It is a highly questionable methodology to assume that if analyses of a sample of (only blind) ascriptions of truth conclude by saying that there are no truth-bearers, then there can be *no* entity capable of possessing truth. Why should we think that the concept of truth is gratuitous simply because we can paraphrase the word ‘true’ away? The most favourable conclusion one can draw, if Williams is correct, is that our language, despite appearances to the contrary, does not presuppose the existence of truth-bearers, *nor* does it eliminate their possibility either.<sup>191</sup> Indeed, Williams’ own theory cannot do without presupposing the existence of truth-bearers. Setting aside the requirement that there is a unique proposition said by Dummett, Williams’ theory for blind ascriptions amounts to the following:

$$\exists p (\text{Dummett says that } p \wedge p)$$

On pages *xii* and 28 of *What is Truth?*, Williams suggests that this is equivalent to:

---

<sup>190</sup> Williams 1979 pp. *xiv* - *xv* points out that it is up to the reader to determine how his theory for blind ascriptions carries over to transparent ascriptions.

<sup>191</sup> Alston puts the same point differently: “Why shouldn’t we say that the equivalence shows that when we assert [an instance of the *T*-scheme] we are really, contrary to surface appearances, attributing the property of truth to a proposition?” (Alston 1996, p. 49).

$\exists p(\text{Dummett's statement states that } p \wedge p)$

which in turn suggests the following equivalent:

$\forall x(x \text{ is true} \leftrightarrow \exists p(x \text{ states that } p \wedge p))$

which is a quantified sentence with  $x$  ranging over statements, a fact Williams denies as belonging to his theory of truth.<sup>192</sup> Connected to this issue is the complaint that no *unified* or general account of truth ascriptions is being offered – only that different types of sentences imply different types of truth claims. More significantly, entities answering to truth-bearers *will* be postulated given that such approaches aim at constructing (adequate) explanatory theories. Kirkham makes this point well:

Our interest begins with a certain value choice: we value having justified beliefs, and that choice makes scepticism a problem, and solving the problem of scepticism leads us to the justification project, and that, in turn, requires a solution to the metaphysical project, at which point the question of what sort of thing can bear truth arises or, more correctly, at which point we *postulate* that certain kinds of entities can be true or false. At no point do we presuppose that there are truth bearers and/or that truth ascriptions refer to them.

(Kirkham 1992, p. 72)

### 4.2.3 Prosententialism

Grover, Camp & Belnap (henceforth G.C.B.) formulate the *Prosentential Theory of Truth* in G.C.B.1975.<sup>193</sup> The key idea underlying this account is that the word ‘true’ has no independent meaning outside of phrases like ‘That is true’ or ‘It is true’ any more than ‘re’ does outside of ‘heresy’ or ‘lu’ outside of ‘mellifluous’. Phrases of the ‘That is true’ ilk are semantically atomic and, thus, the belief in a separable truth predicate is mistaken precisely because the surface grammar of language misdirects us into viewing the predicate as a *genuine* predicate. It is inseparable, they argue, from ‘That is true’ or ‘It is true’ constructions, once we come to understand the deep structure of language.<sup>194</sup> Truth is, therefore, not a predicate nor is it a property-ascribing

---

<sup>192</sup> Kirkham makes a similar observation in Kirkham 1992, p. 322.

<sup>193</sup> See also Grover 1990, 2001 & 2002 for later defences of prosententialism.

<sup>194</sup> This is a very clear departure from (the typical deflationary) endorsement of the *T*-scheme, according to which ‘ $p$  is true’ takes subject-predicate form.

locution. To elucidate the idea, G.C.B. take advantage of the grammatical category *prosentence*. Taking their cue from the redundantist's paraphrase of 'Everything Michael Dummett says is true':

For all  $P$ , if Michael Dummett says  $P$ , then  $P$

G.C.B. argue that we should observe that just as nominal variables have their 'pronoun' counterparts in natural language, so propositional/sentential variables have prosentences as their natural language counterparts.<sup>195</sup> Commonly, pronouns are employed in 'reading off' logical expressions. For example, the sentence ' $\forall x(Fx \rightarrow Gx)$ ' is rendered into natural language as 'Everything is such that if *it* is  $F$ , then *it* is  $G$ ' where the pronoun 'it' stands in for the variable ' $x$ '. Naturally, we cannot similarly render our statement of faith in Michael Dummett's truth-telling ability as we would then get the ill-formed 'Everything is such that if Michael Dummett says it, then it'. What is required to get the comparable paraphrase is an expression:

which is like a pronoun, but which occupies a sentential position. What is wanted is a *prosentence*.  
(G.C.B. 1975, p. 82)

Hence, employing the prosentence 'it is true' we get the rephrase 'Everything is such that if Michael Dummett says it, then it is true', where 'it is true' stands to the propositional variable ' $P$ ' in that same way 'it' stands to the variable ' $x$ ' in ' $\forall x(Fx \rightarrow Gx)$ '.

On pages 86-87, G.C.B. 1975 elucidate on *proforms* – that grammatical category whose members stand in for their anaphoric substituends in the sense that the latter "might just as well" have taken the place of the former (p. 84). Prosentences – as a proform – inherit their content from antecedent statements, just as pronouns inherit their reference from antecedent singular terms. In the following example:

*Bill*: There are people on Mars

*Mary*: *That is true*

'That is true' serves as a prosentence (of laziness, since whatever content 'That is true' has is inherited from what Bill says). Owing to prosentences standing in for other sentences, prosententialists claim that the sentences ' $p$ ' and ' $p$  is true' have the same content. Prosententialists depart company with redundantists here by *not* claiming the word 'true' to be redundant. After all,

---

<sup>195</sup> Redundantism is given a thorough treatment below.

anaphora are important speech acts and ‘it/that is true’ is not eliminable, hence they do not believe we can say without ‘is true’ anything we can with it. They do maintain, however, that sentences containing the word ‘true’ are (in some sense) semantically equivalent to those that do not, i.e. from those that bequeath the content. Clearly, one can think of many more kinds of prosentence. Each of the following is also a prosentence:

Goldbach’s conjecture is true

‘Snow is white’ is true

The claim that penguins waddle is true

where each is formed by conjoining an expression that refers to a sentence to the truth predicate. According to prosententialism about truth, these express no more than the following sentences, respectively:

Every even number is the sum of two primes

Snow is white

Penguins waddle

It would be useful here to briefly point out how prosententialism about truth, as a deflationary theory of truth, participates in the *weak* form of correspondentism outlined in chapter 3. Grover writes:

Many other truth theories assume that a sentence containing a truth predication, e.g., ‘That is true’, is about its antecedent sentence (‘Chicago is large’) or an antecedent proposition. By contrast, the prosentential account is that ‘That is true’ does not say anything about its antecedent sentence (e.g., ‘Chicago is large’) but says something about an extralinguistic subject (e.g., Chicago). (Grover 1992, p. 221)

The truth predicate is not used to say something about truth bearers; in particular, it is not involved in ascribing a substantive property to them. Rather it is employed to say something about the world.<sup>196</sup> When a referring expression is used to explicitly mention an antecedent utterance token – as in ‘Michael Dummett’s last claim is true’ – instead of talking about the utterance, the locutor:

---

<sup>196</sup> Quine 1970 makes exactly the same point in defence of his (deflationary) *disquotationalist* view on truth: the truth predicate serves to “point through the sentence to reality; it serves as a reminder that though sentences are mentioned,

expresses an opinion about whatever (extralinguistic thing) it [is that Michael Dummett] expresse[s]  
an opinion about.  
(Grover 1992, p. 19)

Anaphoric inheritors of content are focused entirely on the world, and not on language.

In addition to proforms of laziness there are proforms that play a *quantificational* role. Consider the following example (taken from G.C.B. 1975, p. 85):

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

The pronoun ‘it’ does not pick up its referent from an antecedent in the same straightforward way as pronouns of laziness do. Replacing ‘it’ by the apparent antecedent ‘each positive integer’ yields the following:

Each positive integer is such that if each positive integer is even, adding 1 to each positive integer yields an odd number

This clearly does not express the sense of the original sentence; ‘each positive integer’ cannot therefore be construed as a referring expression despite the fact that it is the antecedent of ‘it’. Rather, ‘each positive integer’ picks out a family of admissible expressions that can be substituted into the claim. It is therefore to be represented by:

$$\forall x\{(x \text{ is a positive integer} \ \& \ x \text{ is even}) \rightarrow \text{adding 1 to } x \text{ yields an odd number}\}$$

where the (restricted substitutional) quantifier determines the class of substituends. The prosententialist needs to be careful here to ensure that the universe of discourse is always suitably restricted, i.e. not universal. For in the case of ‘Everything Michael Dummett says is true’, with an unrestricted substitutional quantifier, the domain of the sentential variable in the prosententialist’s paraphrase would be all things that can be said, with (indefinitely large) vacuously true conjuncts (as there will be indefinitely many things Michael Dummett does not say rendering the antecedent of these individual conditionals false). There is the further subtlety that, once the relevant domain of

---

*reality is still the whole point*” (pp. 10-11). Similarly, Grover claims that prosentences function “*at the level of the object language*” (*Ibid.*, p. 221).

discourse has been identified, each quantificational prosentence will be tied to a distinct domain without any clue as to what each has in common. For this reason, Belnap 1973 suggests replacing the material conditional with that of a *conditional assertion*. A conditional assertion does not assert the conditional ‘If *A* then *B*’ but rather asserts *B* on the condition that *A*. Hence, in the quantificational prosentence, instead of asserting an infinite conjunction of conditionals, a conjunction of individual (atomic) claims are asserted, all of which satisfy the stipulated condition (in this case, that *x* is a positive integer and an even number).

Prosentences, then, are phrases, which occupy the place of a declarative sentence, possess an antecedent and can be employed in either of the lazy or the quantificational ways. Whenever ‘is true’ occurs outside the context of a prosentence, the surface grammar of language is misleading us into thinking of truth predicatively. G.C.B. (p. 96) diagnose this state of affairs by saying that owing to the frequency with which so many expressions in language take the subject-predicate form it seems altogether convenient to maintain that paradigm.

There are a number of objections one can raise against the prosententialist thesis:

1. The argument that ‘true’ semantically functions in the same way that proforms do without telling us what proforms – such as ‘he’, ‘it’, ‘so’ etc. – have in common with ‘it is true’ and ‘that is true’ is no argument at all. All that has been identified is an idle analogy. What is required is a systematic answer to the question: *in virtue of what do ‘it is true’ and ‘that is true’ work as proforms?* Worse, it does not follow that the word ‘true’ is *essentially proformal* once an analogy between the paraphrases of ‘Everything Michael Dummett says is true’ and, say, ‘Everything green is coloured’ have been identified. Respectively, the paraphrases are ‘Everything is such that if Michael Dummett says it is true, then it is true’ and ‘Everything is such that if it is green, then it is coloured’ (where the quantifier in each determines the class of anaphoric substituends). The semantic effects of ‘true’ – such as the inferential relationship between truth predicated sentences and non-truth predicated sentences – can be described and explained *without* making any appeal to proformal similarities.
2. The locution ‘That is true’ need not always serve as a prosentence of laziness. Consider Bill’s saying ‘There are people on Mars’ and Mary responding likewise: ‘There are people on Mars’. This is not the same as affirming what Bill said, since Mary might just like

repeating Bill's locutions independently of whether she believes in the content of what he expresses.

3. According to G.C.B., each occurrence of 'it' and 'that' in a prosentence is not to be interpreted as a referring expression with an independent meaning; 'it is true' and 'that is true' are semantically atomic. This is part of their case against the claim that truth is a separable predicate engaged in property-ascribing locutions. Hence, G.C.B. are forced to find prosentences where there don't seem to be any, long-windedly paraphrasing (what appear to be) truth predications by typically employing quantification to preserve prosentences like "it is true" as an unbroken unit. They argue that, despite appearances to the contrary, "'Snow is white' is true" is not composed of the nominalising, referring expression 'Snow is white' and the (separable) truth predicate '...is true' but has instead the true logical form "For any sentence, if it is 'Snow is white', then it is true" or 'Consider: snow is white. That is true'. But there seems to be no motivation for supposing that "'Snow is white' is true" hides genuine quantifiers underneath. Does it not seem equally plausible to suppose that constructions like "it is true" contain pronouns that refer to some previously mentioned truth bearer to which the property of truth is being attributed? Moreover, there are a whole range of inferences that involve truth which equipollently suggest that truth indeed *is* a predicate. Paul Horwich is keen to make this point. When we infer "from  $x = \textit{that } p$ ' and ' $x$  is true' to ' $\textit{that } p$  is true', and hence to ' $p$ '",<sup>197</sup> such inferences are most readily made sense of by supposing that truth functions as a predicate.
4. Robert Brandom 1994 defends the prosententialist account of truth, but (perhaps for the objection just alluded to) does not view "it is true" and "that is true" as atomic. He insists rather that '...is true' should be treated as a prosentence-forming operator in that when conjoined with a nominalised sentence or referring expression the result has the same content as the sentence nominalised or expression referred to. However, as Wilson 1990 points out, nominalised sentences and referring expressions do not always behave like proper names – they can bring with them more baggage than is necessary to succeed in referring.<sup>198</sup> Consider the following example:

*Bill:* John has just been awarded the Lakatos Scholarship

---

<sup>197</sup> Horwich 1998a, p. 125.

<sup>198</sup> See Wilson 1990, p. 26.

Mary: What that good for nothing so-and-so just said came true

Mary's utterance, if the prosententialist is right, should have that same content as Bill's. But it is clear that Mary's remark does more than reassert the content of what Bill is saying.

5. As will be amplified below in describing redundantism's central weakness, in arguing that truth is not a property expressed by the truth predicate (because there is *no* such genuine truth predicate), if the analysis employs sentential variables with quantifiers to bind them, the wherewithal to characterize exactly the property that all truths share become available after all. It then becomes difficult to maintain that '...is true' is not a predicate applicable to exactly those things that are true.

### 4.3 *T-scheme Based Deflationism*

There are three principal camps into which *T*-scheme based deflationists fall: *redundantism*, *disquotationalism* and *minimalism*. We first describe the tenets of redundantism before moving on to identifying its central weakness. Disquotationalism and minimalism will then subsequently be given the same treatment.

#### 4.3.1 *Redundantism*

Central to the redundantist's claim is that, as an example, the proposition expressed by (1) below is to be considered the same as the proposition expressed by (2):

(1) The thought that 5 is a prime number is true

(2) 5 is a prime number

This particular illustration is taken from Gottlob Frege. This is how he expressed his 'equivalence' view:

One can, indeed, say: 'The thought, that 5 is a prime number, is true'. But closer examination shows that no 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 a 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'.  
(Frege 1892 (1993), p. 30)



Each of the redundantists to be considered subscribes to the view that truth is a redundant predicate in the fashion just adumbrated. The reasons proffered, however, differ across them. There are in addition two types of class into which redundantists can be classified: ‘no property’ redundandists – who believe that there is simply *no property* of truth answering to this redundant predicate – and ‘property’ redundantists – who believe otherwise, that truth is in fact a property (albeit a redundant one). Frege is the most famous exponent of ‘no property’ redundantism;<sup>199</sup> Ramsey most famous for ‘property’ redundantism. We begin by continuing our exposition of Frege.

Frege was an actually an *indefinabilist* about truth.<sup>200</sup> truth, according to him, is a *sui generis* notion not amenable to being reduced to explanatorily more primitive notions:

7. What true is, I hold to be indefinable...

16. Following the laws of logic can guarantee the truth of a judgement only insofar as our original grounds for making it, reside in judgements that are true,  
(Frege 1906, or earlier (1979), pp. 174-175)

In fact, Frege went even further claiming that truth is *primordial* – we cannot even think of a proposition without presupposing that it is true:

...we cannot recognize a property of a thing without at the same time realizing the thought that this thing has this property to be true. So with every property of a thing there is joined a property of a thought, namely, that of truth.  
(Frege 1911 (1999), p. 88)

Frege’s view that truth is a simple, unanalyzable quality is intimately tied in with his systematic considerations involving sense and reference.<sup>201</sup> According to Frege, the sense of a statement is distinct from its reference. For example, ‘the *Stagirian* teacher of Alexander the Great’ and ‘the founder of formal logic’ have different senses despite both having the same reference – Aristotle. Guided by the principle that the sense (or reference) of a complex expression is a function of the sense (or reference) of the atoms composing it, Frege observed that in a large number of cases the substitution of co-referring expressions with inequivalent senses alters the proposition expressed,

<sup>199</sup> A.J. Ayer is also considered a central figure in maintaining the twin theses of the truth predicate’s *redundancy* and the claim that there is *no property* of truth answering to this redundant predicate (see Ayer 1963, 1953 and 1963). However, a commentary on his works relevant to his deflationism *in addition* to that supplied of Frege’s would be redundant: for the purposes of this background chapter, the lessons to be learnt from both are equivalent.

<sup>200</sup> Hans Sluga in Sluga 2002 provides an excellent overview of, and background to, Frege’s indefinabilism about truth.

<sup>201</sup> Outlined in Frege’s “Über Sinn und Bedeutung” (1892).

whereas the same is not true for substitutions of synonymous expressions. Since he also observed that the only thing that remains unaltered for expressions in which co-referring atoms have been substituted is their truth-value, Frege was led to identifying the proposition expressed by a statement with its sense and the truth-value with its referent. As referents, truth-values are only to be conceived of as objects and so can never form part of a thought; they cannot be the sorts of things that are the constituents of propositions. In particular, they cannot be the sorts of things that are the constituents of propositions expressed by statements containing the truth predicate. This led Frege to maintain that the truth predicate is not a genuine predicate (it is not used to describe anything)<sup>202</sup> and that, therefore, the proposition expressed by (1) above is the same as the proposition expressed by (2)

Many will find the premises of his argument concerning the sense and reference of statements unacceptable.<sup>203</sup> But, as Scott Soames observes in *Understanding Truth*,<sup>204</sup> Frege's argument fails to succeed *even if* we were to accept Frege's views on sense and reference. Given that the *True* is a referent determined by true thoughts, all one needs to do to define the truth predicate is assert the following:

*x* is true if and only if *x* is a thought that determines the *True* as referent

Hence, the thought expressed by 'the thought that 5 is a prime number is true' is the one that predicates the property of truth so defined to the (different) thought expressed by '5 is a prime number'.

Frege also articulated what is understood to be another central deflationary theme: that there are two different types of 'linguistic environments'<sup>205</sup> in which the truth predicate is used:

It is...worthy of notice that the sentence 'I smell the scent of violets' has just the same content as the sentence '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 the scientist after much hesitation and careful inquiry, can finally say 'what I supposed is true'? The meaning of the word 'true' seems to be altogether unique.  
(Frege 1911 (1999), p. 88)

---

<sup>202</sup> This view is referred to by Alfred Tarski as "*the nihilistic approach to the theory of truth*" (Tarski 1969, p.66).

<sup>203</sup> In the case of singular definite descriptions, for example, Russell's theory of descriptions is preferred.

<sup>204</sup> Soames 1999, p.47.

<sup>205</sup> I use Soames' terminology from Soames 1997, p. 2.

We can identify the two environments as:

*Linguistic Environment A*: It is true that *S* / The proposition that *S* is true

*Linguistic Environment B*: Everything Michael says is true

A valid argument is one where if the premises are true then so is the conclusion

All consequences of true statements are true<sup>206</sup>

Frege observed that uses of the truth predicate in environment *A* – such as in ‘It is true that I smell the scent of violets’ – signal the predicate’s redundancy in that nothing of content is being added to the proposition ‘I smell the scent of violets’: they express the same thought (though, as indicated above, Frege’s own way of justifying this is defeasible). One might tentatively say that ‘It is true that *S*’ and ‘*S*’ are trivially equivalent, and that such environments are important in identifying what the concept of truth consists in. Frege failed, however, to investigate what might be of significance for uses of the truth predicate in environment *B* – blind ascriptions of truth, as in ‘what I supposed is true’ – but it is abundantly clear that truth is not easily eliminable in such environments without loss in content: the proposition of which truth is predicated is not explicitly displayed and the structure of the sentence is lost. It is in such circumstances, redundantists argue, that the truth predicate derives its utility. We employ the concept of truth precisely to accomplish such linguistic tasks.

Frank Ramsey is also famously regarded as a redundantist about truth. However, Ramsey’s path to redundantism differs quite starkly from Frege’s. According to Ramsey, we need to sharply distinguish (in contradistinction to many traditional accounts of ‘substantive’ truth) theories of truth from theories of content (or meaning). Once this has been achieved and a theory of content is determined, we ought to see then that:

there is really no separate problem of truth but merely a linguistic muddle.  
(Ramsey 1927 (1964), p. 16)

For Ramsey (and contrary to what Frege thought) truth *is* a property, and in particular a property of propositions, which themselves supply truth-conditions, thereby rendering redundant any truth ascriptions one might make. Hence, for Ramsey, “the real problem is not as to the nature of truth and falsehood, but as to the nature of judgement or assertion” (*Ibid.*).

---

<sup>206</sup> Cf. Soames 1999, pp. 2-3.

Ramsey claimed that all uses of the word ‘true’ are, in principle, eliminable. This is due to the fact that (to employ the example he is most famous for using) “‘Caesar was murdered’ is true” means no more than Caesar was murdered.<sup>207</sup> Hence, “‘Caesar was murdered’ is true” does not say anything about the proposition ‘Caesar was murdered’, but rather it says something about Caesar (namely, that he was murdered). In cases of explicit mention, an attribution of truth to the explicitly given sentence can always be replaced by the use of that very same sentence: the same content is being expressed when one asserts a proposition as when one asserts that that proposition is true (and *mutatis mutandis* for falsity ascriptions).

It is clear, from our discussion of Frege, that not all uses of the truth predicate are so easily eliminable and, thus, not wholly redundant. Ramsey appreciated this (and, contrary to Frege, went on to investigate its significance). Truth may also be attributed to propositions that are not explicitly given. I may not know what Michael Dummett has to say about Oxford academics, but, having the utmost faith in his honesty, knowledge and sagacity, I may still insist that, no matter what he has to say about them, it is true. Other examples of attributions of truth to propositions merely described or indicated (and thus employ the truth predicate in a non-easily eliminable manner) involve generalizations. Consider the following example:

Everything Michael Dummett says is true

This means simply that the propositions asserted by Michael Dummett are true. The truth predicate cannot, on pain of being ungrammatical, be so easily eliminated:

In the...case in which the proposition is described and not given explicitly...we get statements from which we cannot in ordinary language eliminate the words ‘true’ and ‘false’. Thus if I say ‘He is always right’, I mean that the propositions he asserts are always true, and there does not seem to be any way of expressing this without using the word ‘true’.<sup>208</sup>  
(Ramsey 1927 (1964), p. 17)

---

<sup>207</sup> *Ibid.* p. 17. Ramsey’s sentence actually reads like this: “*Its value ‘Caesar was murdered is true’ is the same as ‘Caesar was murdered’*” which fails to convey precisely what Ramsey had in mind, since *Caesar was murdered* cannot serve as the subject of a sentence without being ensconced within quotation marks.

<sup>208</sup> Without employing the truth predicate, we would get the following kind of ungrammatical, ill-formed instances:

If he says the proposition that the law of excluded middle fails, then the proposition that the law of excluded middle fails

However, according to Ramsey, the case for truth's redundancy can *still* be made: even in such blind ascriptions of truth, the concept can be quite straightforwardly paraphrased away. For the example Ramsey cited, 'He is always right' can be paraphrased thus:

For all  $p$ , if he asserts  $p$ , then  $p$

But this proposal is problematic.<sup>209</sup> As the quantification over the propositions involved is first order, the variable ' $p$ ' can stand only as a substituent for names or singular terms that refer to propositions; the second occurrence of ' $p$ ' cannot, therefore, occupy – as it is doing – the position of a sentence. One proposal out of this, which might be available to the redundantist, is to understand the quantifier involved as a *higher order* type of quantification.<sup>210</sup> Let us consider the employment of higher order quantification, as in:

(\*) For all  $P$ , if he asserts  $P$ , then  $P$

where the variable ' $P$ ' ranges over sentences – instead of names or singular terms – thereby permitting us to quantify into sentence position. We obtain instances of (\*) by replacing variables with propositions, as in:

If he asserts that the law of excluded middle fails, then the law of excluded middle fails

Such a construal is clearly a way out of the syntactical inadequacies Ramsey talked of. However, it does oblige the redundantist to explain *how* one is to understand the higher order quantifier involved. What we are *not* after here is how one might *define* the quantifier; it is standard for any such definition to spell out the quantified sentence's truth conditions and thus employ the notion of truth. Rather, the question is: 'What is the nature of (what should be understood as) the 'primitive'

---

<sup>209</sup> As highlighted in numerous places, for example in Soames 1999, Künne 2003, Stoltjar, D. & Damnjanovic, N. 2007.

<sup>210</sup> In a paper published posthumously, Ramsey was remarkably prescient in identifying a number of deflationary themes that currently occupy centre-stage. In addition to suggesting the equivalence of 'the proposition that  $p$ ' and ' $p$ ', he also presented a contemporary explanation of why this fails to imply the total redundancy of the English truth predicate – something that anticipates prosententialism about truth': "*As we claim to have defined truth we ought to be able to substitute our definition for the word 'true' whenever it occurs. But the difficulty we have mentioned renders this impossible in ordinary language which treats what really should be called pro-sentences as if they were pronouns. The only pro-sentences admitted by ordinary language are 'yes' and 'no', which are regarded as by themselves expressing a complete sense, whereas 'that' and 'what' even when functioning as short for sentences always require to be supplied with a verb: this verb is often 'is true' and this peculiarity of language gives rise to artificial problems as to the nature of truth, which disappear at once when they are expressed in logical symbolism, in which we can render "what he believed is true" by "if  $p$  was what he believed,  $p$ ".*" (Ramsey 1991 (2001), p. 438).

quantifier that participates in the redundancy *analysis* of truth?’<sup>211</sup> One option is to understand the quantifier substitutionally. Under this interpretation, the quantifier in (\*) is associated with a ‘substitution class’ – a class of expressions that can be substituted for the variable ‘*P*’ bound by the quantifier. A ‘substitution instance’ is obtained by removing the quantifier and substituting for the bound variable a member from this class. Hence, the universally quantified sentence (\*) in this account, is true just in case each substitution instance is true. The first cause of disagreement with this option is that it is clearly parasitic on some (other) antecedent notion of truth. The second is that it is highly limitative: the sentences in the substitution class may be inadequate to express *every* proposition. The language in which (\*) is couched is English, so its associated substitution class is the class of (the logical forms of) English sentences, and English does not have the resources to express all and every proposition. There may be propositions about objects unnamed by proper names in English and about properties not expressed by English predicates that speakers of the language containing such resources would be able to express. There are, thus, some propositions relevant to the truth of (\*) that are not captured by its substitution class. Hence, the redundantist cannot advance the higher order substitutional interpretation of (\*) as the correct redundancy analysis of it.<sup>212</sup>

This leaves the option of interpreting the quantifier objectually. Under this interpretation, the quantifier is bound to a domain of objects over which it quantifies. The universally quantified sentence (\*) in this account is true just in case the sentence obtained by erasing the quantifier is true for all assignments of objects in the range of ‘*P*’ to the variable. I will now go on to claim that this position is untenable.

#### 4.3.2 *The Central Objection to Redundantism*

In our discussion of Ramsey, we concluded that the deflationary truth-eliminating paraphrasis of blind truth ascriptions could not be accomplished without the employment of higher order quantification that range over sentences. There is no way of achieving such a paraphrasis, and still engage in what might properly be called a deflationary theory, without the employment of such an apparatus. For example, in:

Everything Michael Dummett says is true

---

<sup>211</sup> Cf. Soames 1997, p. 7.

<sup>212</sup> There are further problems connected with propositions expressible by different sentences at difference times.

we get the following deflationary paraphrase:

For all  $P$ , if Michael Dummett says  $P$ , then  $P$

where ' $P$ ' is a sentential variable. Naturally, the redundantist is obliged to spell out how the quantifier is to be understood. In our examination of the substitutional option, we concluded that that there can be no redundancy analysis involving higher order substitutional quantification – the substitution class is inherently limited in (potentially) not having any sentence that expresses something/s Michael Dummett says. Also, such quantification would presuppose the concept it is being invoked to replace. This leaves the option of interpreting the higher order quantifier objectually. There is an immediate problem for redundantism with this. Such quantification is ontologically committed to the entities it quantifies over. If it quantifies over propositions, say, then this would undermine the programme (central to people like Williams and Strawson) of doing away with truth bearers. Furthermore, once the formal apparatus demonstrating truth's redundancy is in place, the redundantist would be hard pressed to deny that there is *indeed* a property associated with it. One could characterize that property shared by all truth bearers in the following way:

For all objects  $x$  [ $x$  is true iff  $\exists P(x = \text{the proposition that } P \text{ and } P)$ ]

Alternatively, if the redundantist wishes to deny the existence of propositions and give instead a theory of sentential truth, then the property of truth can be determined thus:

For all objects  $x$  [ $x$  is true iff  $\exists S(x \text{ expresses } S \text{ and } S)$ ]

The point is that the claim of the truth predicate's redundancy rests on the provision of a coherently motivated type of quantification. The redundantist is committed to some principle of the above form. Once this has been achieved:

- (1) the means by which to characterize that property shared by all truths become available,  
and
- (2) it becomes impossible to shy away from a commitment to the existence of truth bearers.

### 4.3.3 Disquotationalism

The central defining disquotationalist thesis is this:

*The truth predicate is a logico-mathematical device of disquotation serving the purpose of (finitarily) expressing infinite logical conjunctions/disjunctions*

Both Quine and Field are regarded as the foremost defenders of this position. For our purposes, we focus exclusively on Quine.<sup>213</sup> Quine articulates his views in a number of places (see, for example, Quine 1970, 1974b, 1992 and 1995). According to Quine, once sentences of one's own language have been identified as the appropriate bearers of truth, then - once names of these sentences are obtained by placing quotation marks around them - the truth predicate may be used for cancelling out these marks.<sup>214</sup> Moreover, since "no sentence is true but reality makes it so"<sup>215</sup> it is owing to the fact that reality is our focus that the truth predicate is endowed with this disquotational feature.

Quine was not a full-blooded correspondentist, seeing as he did in the employment of terms like 'fact' merely what he calls a "put-up job".<sup>216</sup> His endorsement of the correspondence theory is very limited, going only as far as agreeing that "to ascribe truth to the sentence ['Snow is white'] is to ascribe whiteness to snow".<sup>217</sup> Disquotationalism, then, is a form of correspondentism about truth (*weak* correspondentism) but what it insists on, however, is the provision of an *austere* account of truth that does not evoke substantially loaded notions like 'fact' and 'correspondence' while still subserving the correspondence intuition. What one requires, so this deflationist insists, is to perform a kind of 'deflationary surgery' on the correspondence theory in order to eviscerate its metaphysical baggage which can be excised without loss. In considering, for example, the sentence 'Snow is white', then according to correspondentism:

---

<sup>213</sup> Field articulates his particular type of disquotationalism in Field 1986, 1988, 1994a and 1994b.

<sup>214</sup> This might be called the 'disquotation effect': we are permitted to eliminate the truth predicate in predications containing quotations, by inferring from a truth-predication (say, '*snow is white*' is true) to the quoted sentence itself (to *snow is white*). So, according to this view, the truth predicate behaves very much like a logical concept, with rules analogous to those for disjunction and conjunction (rules like, for any sentences A and B,  $A \vdash A \vee B$ , or, for any sentences A and B,  $A \wedge B \vdash A$ ). As Davidson has remarked: "...Tarski's methods allow us to replace the truth predicates he defines in any context, and the replacement leaves no explicitly semantical predicates in its wake; in this respect, his truth predicates are like the sentential connective "it is true that", which may be removed by simple deletion." (Davidson 1990, p.285).

<sup>215</sup> Quine 1970 (1986), p. 19.

<sup>216</sup> Quine 1992, p. 80. See also Quine 1987, p. 213.

<sup>217</sup> Quine 1992, p. 80.



The sentence ‘Snow is white’ is true just in case the sentence corresponds to the fact that snow is white

However, presumably:

The sentence ‘Snow is white’ corresponds to a fact just in case snow is white

which is made true by the ‘whiteness of snow’. This prompts the following shave:

The sentence ‘Snow is white’ is true just in case snow is white

Thus, to say that the sentence ‘Brutus killed Caesar’ is true, or that ‘The atomic weight of sodium is 23’ is true, is, in effect, to say that Brutus killed Caesar, or that the atomic weight of sodium is 23.<sup>218</sup> We may wish to represent this thesis in the following schematic fashion:

*DS*: ‘*p*’ is a true sentence if and only if *p*

*DS* is a schema postulated to apply to (quoted on the left, non-quoted on the right) sentences for some object language which is not necessarily identical to the language in which the equivalences themselves are formulated.<sup>219</sup> *DS* tells us nothing about the *nature* of truth: an analysis of a presumed nature of truth would take the form of stating what that property is that is commonly shared by all and only all true sentences and this schema does not do that. But it is clear what each instance does say once the object language is located. It is also clear that *DS* is a variant of Tarski’s Convention **T** and says something metalinguistic, *viz.* that a materially adequate theory of truth ought to derive all sentences of a particular syntactic form. *DS* is a scheme which is just as hard – or rather just as unproblematic – to comprehend as logical axiom schemes or logical rules, like *modus ponens*.<sup>220</sup> But *DS* leads us into the liar paradox. Consider the self-referring sentence  $\lambda$ :

---

<sup>218</sup> These famous examples come from Quine 1960, p. 24.

<sup>219</sup> The object language is required to be a subset of the metalanguage, for otherwise non-homophonic translations of the object language into the metalanguage would be needed, and hence more than simply disquotation is required.

<sup>220</sup> Marian David argues that *DS*’s intelligibility forces us to invoke more than “*minimal conceptual resources*” (1994, p. 66). This issue of deflationary theories of truth calling upon minimal conceptual resources was discussed earlier in *Introductory Remarks*..

$\lambda$ :  $\lambda$  is not a true sentence

By *DS*, we get:

‘ $\lambda$ ’ is a true sentence if and only if  $\lambda$  is not true

Contradiction! One might argue that So, *DS* is in need of revision. Quine did not consider this implication from the liar. However, Tarski did. Motivated principally by the fact that not all truth ascriptions are to quotation names of sentences, Tarski considered the following revision:

*DS\**:  $x$  is a true sentence if and only if, for some  $p$ ,  $x$  is identical to ‘ $p$ ’, and  $p$

taking advantage of the fact that every sentence has its own quotation name, as expressed by the following principle:<sup>221</sup>

For all  $x$ , if  $x$  is a true sentence, then, for some  $p$ ,  $x$  is identical with ‘ $p$ ’

Clearly, *DS\** helps solve the issue concerning the fact that *DS* is not eliminative in only dealing with a restricted class of sentences. In *DS\**, any singular term denoting a sentence – such as a quotation name or definite description – can be a substituent for ‘ $x$ ’. Despite this, and apart from the fact that *DS\** leads us to the liar paradox,<sup>222</sup> there are intrinsic complications involved with the quantification in *DS\**’s associated universal generalization:

For all  $x$ ,  $x$  is a true sentence if and only if, for some  $p$ ,  $x$  is identical to ‘ $p$ ’, and  $p$

These problems are similar to the quantificational ones plaguing the redundantist. Here, we cannot objectually quantify into quotation marks. Quotational contexts are *referentially opaque*. For example, we cannot objectually quantify into ‘Giorgione’ in “Giorgione is called ‘Giorgione’ because of his size” as the second occurrence of ‘Giorgione’ does not refer to the person

---

<sup>221</sup> Tarski 1936a (1983), p. 159, footnote 1.

<sup>222</sup> David 1994, p. 162 has all the details. It is for this reason that Tarski prefers the *structural-descriptive* method of referring to expressions; furthermore, *DS\** is not something Tarski would endorse since propositional quantification of the form “for some  $p$ ” where “ $p$ ” occurs in sentence positions are unacceptable to him. Issues surrounding the liar paradox are orthogonal to the central concern of this thesis, *viz.* to demonstrate how an authentic correspondence theory can be defended with respect to a model language sharing salient features with a natural language. Moreover, the correspondence theory is itself similarly plagued by semantic paradoxes. But this choice in exposition is fairly standard and leads us quite naturally to the central point here concerning quantification.

Giorgione.<sup>223</sup> The solution would be to substitute *sentences*, and not names of sentences, for the variable ‘*p*’ – something objectual quantification does not allow.

What about substitutional quantification? Standardly, substitutional variables are associated with a substitution class of admissible substituends.<sup>224</sup> A substitutional (particular) quantifier is true just in case the open sentence that obtains from erasing the quantifier has one true substitution instance; a substitutional (universal) quantifier is true just in case every substitution instance of the open sentence that obtains from erasing the quantifier is true. The problems concerning quantification into quotation contexts can thus be eliminated. The disquotationalist may then seek salvage in stipulating that the *definiens* of *DS\**’s universal generalization be governed by the substitutional (particular) quantifier, as in:

$$\forall x\{x \text{ is a true sentence just in case } \exists p(x = \text{‘}p\text{’ and } p)\}$$

The semantics of substitutional quantifiers that govern formulas involving both substitutional and objectual variables would then have to be modified to make any sense at all. The modification required would be to incorporate how the object that is substituted for the variable ‘*x*’ *satisfies* the sentence.<sup>225</sup> But herein lies the problem. Not only does pure substitutional quantification presuppose that the notion of truth is already at hand – it is explained in terms of the *truth* of its substitution instances – the semantics of substitutional quantification over mixed (substitutional and objectual) variables would have to employ the highly substantial, non-deflationary tool of *satisfaction*. The disquotationalist would not want to avail herself of this particular resource.

So far, Quine asked us to appreciate that an attribution of truth to a sentence merely undoes the effects of the quotation used to nominalize the sentence. “In speaking of the truth of a given sentence”, Quine writes:

...there is only indirection; we do better simply to say the sentence and so speak not about language but about the world.  
(Quine 1970 (1986), p. 11)

---

<sup>223</sup> This is taken from Quine 1953a (1980), p. 140.

<sup>224</sup> See Kripke 1976.

<sup>225</sup> This can be easily done – see David 1994, p. 84 for the details.

Why then, if the truth predicate is superfluous, did Quine think we have a truth predicate at all? Consider the following: from the sentences ‘Socrates is mortal’, ‘Aristotle is mortal’, ‘Aristophanes is mortal’, etc. one can generalize to ‘For all  $x$ , if  $x$  is a man, then  $x$  is mortal’. However, there are problematic cases where a similar move cannot be made if the quantifier is read objectually. For, consider ‘If time flies then time flies’, ‘If Bob is hungry then Bob is hungry’, etc. On the model given, our generalization would be the incoherent ‘For all  $p$ , if  $p$  then  $p$ ’ – we would want the variable appearing in the consequent to play a sentential role, something which is incompatible with the objectual reading of the quantifier. The truth predicate comes to the rescue by allowing us to make the required generalization coherently by turning a sentence nominalization into something the use of which is equivalent to a use of that sentence; truth acts as a ‘de-nominalizer’:

We could not generalize as in ‘All men are mortal’, because ‘time flies’ is not, like ‘Socrates’, a name of one of a range of objects (men) over which to generalize. We cleared this obstacle by *semantic ascent*: by ascending to a level where there were indeed objects over which to generalize, namely linguistic objects, sentences.  
(Quine 1992, p. 81)

By ‘semantically ascending’, our discourse is about language – sentences that share a common feature. We can then restore the structure of a complete sentence.<sup>226</sup> The truth predicate allows us to descend – cancelling the effect of semantic ascent – and talk instead of the world:<sup>227</sup>

This ascent to a linguistic plane of reference is only a momentary retreat from the world, for the utility of the truth predicate is precisely the cancellation of linguistic reference. The truth predicate is a reminder that, despite a technical ascent to talk of sentences, our eye is on the world. This cancellatory force of the truth predicate is explicit in Tarski's paradigm: ‘Snow is white’ is true if and only if snow is white. Quotation marks make all the difference between talking about words and taking about snow...By calling the sentence true we call snow white. The truth predicate is a device for disquotation.  
(Quine 1970 (1986), p. 12)

In principle then, the truth predicate can be eliminated by replacing an open formula such as:

---

<sup>226</sup> Cf. Soames 1984, p. 412.

<sup>227</sup> We could represent the idea that the truth predicate allows one to semantically ascend/descend as *syntactical operations*:

	<u>Semantic ascent:</u>	
From the premise:	snow is white	$S$
You infer:	‘snow is white’ is true	‘ $S$ is true’
	<u>Semantic descent (disquotation):</u>	
From the premise:	‘snow is white’ is true	‘ $S$ is true’
You may infer:	snow is white	$S$

(a)  $x$  is true

in favour of the infinite disjunction:

(b) ( $x =$  ‘snow is white’ and snow is white) or ( $x =$  ‘grass is green’ and grass is green)  
or ( $x =$  ‘penguins waddle’ and penguins waddle) or ...

or the infinite conjunction:

(c) (If  $x$  is “snow is white” then snow is white) & (if  $x$  is “grass is green” then grass is green) &  
(if  $x$  is “The Moon is made of cheese” then the Moon is made of cheese) & ...

The argument is that the truth predicate ‘serves this purpose’ of “expressing infinite conjunctions and disjunctions”.<sup>228</sup> Note that this purpose also covers blind ascriptions of truth – both the purpose of expressing generalizations and that of blind ascriptions are actually intimately connected.

Consider the blind ascription:

The sentence most favoured by Tarski is true

which is equivalent to the following infinite conjunction:

If the sentence most favoured by Tarski is ‘Snow is white’ then snow is white, and if the sentence most favoured by Tarski is ‘The elementary theory of groups is undecidable’,<sup>229</sup> then the elementary theory of groups is undecidable, and if...

To the extent that the truth predicate has any value (as it has no metaphysical depth), it increases our expressive power by serving the (sole) purpose of expressing infinite conjunctions and disjunctions:

---

<sup>228</sup> Similarly, following Quine’s lead, Stephen Leeds writes, “*It is not surprising that we should have use for a predicate  $P$  with the property that “‘...’ is  $P$ ” and ‘...’ are always interdeducible. For we frequently find ourselves in a position to assert each sentence in a certain infinite set  $z$  (e.g., when all the members of  $z$  share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of  $z$  is warranted. A predicate  $P$  with the property described allows us to construct such a sentence:  $\forall x (x \in z \rightarrow P(x))$ . Truth is thus a notion that we might reasonably want to have at hand, for expressing semantic ascent and descent, infinite conjunction and disjunction.*” (Leeds 1978, p. 121)

<sup>229</sup> This particular example of what Tarski’s most favoured sentence might have been is borrowed from Volker Halbach (Halbach 2001a, p. 174).

We affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have restored to semantic ascent.  
(Quine 1970 (1986), p. 12)

Though Quine is referring here only to universal generalizations, the point clearly carries over to existential generalizations as well. This particular quotation is, in fact, referred to by Volker Halbach as:

serv[ing] as a motto for many disquotationalist accounts [after Quine].  
(Halbach 2004, p. 66).

Most deflationists tend to append to the above insights the following remark: once we have spotted its logical features and understood the use to which truth is put, this is *all* there is to be said about the concept.<sup>230</sup>

Since the central objections to be raised against disquotationalism are of equal force against minimalism, we turn to a critical exposition of minimalism first before coming to these objections.

#### 4.3.4 *Minimalism*

Paul Horwich's *Minimalist Theory of Truth* takes propositions to be the appropriate bearers of truth. Letting  $\langle p \rangle$  stand for "the proposition that  $p$ " – a term-forming operator which, when applied to the sentence  $p$ , yields the singular term referring to the proposition expressed by  $p$  – the following equivalence schema lies at the bedrock of Horwich's theory:

$$(ES) \quad \langle p \rangle \text{ is true iff } p$$

Schema (ES) is understood to express a propositional function. It takes any proposition  $p$  into the result of applying the schema to that proposition:

[W]hen applied to any proposition,  $y$ , this structure (or function) yields a corresponding axiom of the minimal theory, MT. In other words, the axioms of MT are given by the principle

For any object  $x$ :  $x$  is an axiom of the minimal theory if and only if, for some  $y$ , when the function  $E^*$  is

---

<sup>230</sup> This is the *completeness condition* in the following chapter.

applied to  $y$ , its value is  $x$ .

Or in logical notation

$\forall x(x \text{ is an axiom of MT} \leftrightarrow (\exists y)(x = \text{ES}(y)))$ .  
(Horwich 1998a, pp 19-20)

In *The Structure and Content of Truth* (Davidson 1996), Davidson raised the following question related to the schematic generalization (ES): instances are obtained by replacing ' $p$ ' occurrences by tokens of English sentences (or tokens of sentences from possible extensions of English) and both are placed in *different contexts* (the first forms part of the referring term  $\langle p \rangle$ ). How, then, are these two connected? Horwich 1998b replies by arguing that certain conditions need to be placed on such instantiations. These are:

- (a) each ' $p$ ' is to be replaced by tokens of an actual or possible English sentence,<sup>231</sup>
- (b) these tokens are given the same interpretation,
- (c) under that interpretation they express a proposition, and
- (d) 'that' and 'proposition' terms are given their English meanings.

This means that when specifying the axioms of the minimalist theory of truth, we must employ semantic notions such as 'language', the relation of 'expressing', 'interpretation' and 'proposition'. Minimalism relies on these heavily but, clearly, the truth-concept cannot be used to explain these. The whole challenge is to explain the concept of truth non-circularly. Horwich evokes the Wittgensteinian account of *meaning* – that use exhaustively determines meaning – to help overcome this; he argues that a use-theoretic account could be given of, for example, " $u$  and  $v$  express the same proposition".<sup>232</sup> However, a dilemma still lingers. Consider "this is a cat". An account of its meaning could proceed thus: '*this*' is an indexical *referring* to an associated object on occasions of use; '*cat*' is a *true* predicate of all and only cats. It is easy to see that the desire for non-circular reasoning quickly evaporates here – semantic notions, such as reference and truth, are presupposed by the account of meaning just given. But let us suppose, instead, that this isn't the case and the account uses notions like causal links to the brain or fixation of attention on things in understanding

---

<sup>231</sup> So Horwich's minimal theory is composed of infinitely many axioms, infinitely many of which we cannot formulate or understand. But note that our understanding of the concept of truth, according to Horwich, consists in our disposition to accept *a priori* those axioms we can formulate and this disposition provides the best explanation of our overall use of the term 'true'.

<sup>232</sup> In fact, minimalists rely heavily upon the Wittgensteinian (use-theoretic) account of meaning. Given their maintenance that it is our disposition to accept all instances of the equivalence schema which constitutes our understanding of the concept of truth, it is only by making an appeal to this account that minimalists can sustain the thesis the meaning of 'true' is *indeed* constituted by this disposition.

our use of demonstratives. To be a successful attempt, wouldn't these have to fill in the notions conjured up to a reasonable approximation to 'reference' and 'truth' in any case? An account of meaning in terms of use can only be satisfactory if it enables us to infer a corresponding *T*-sentence, but such an account would run contrary to the enterprise of deflating notions like reference and truth. As Simon Blackburn and Keith Simmons put it:

They seem to have proved their robustness [of notions like reference and truth] via the robust, non-deflationary story that started out trying to do without them, but ended up constructing them.  
(Blackburn & Simmons 1999, p. 22)

There is, in fact, a dilemma facing the whole deflationist project concerning the choice between dressing up 'sententially' or 'propositionally'. This is discussed in Jackson, Oppy and Smith 1994. The objection is that if deflationists present their theory of truth in accordance with sententialism then it is simply false; if, on the other hand, they present their theory in accordance with propositionalism then it is trivial. Consider the following claim:

'Snow is white' is true if and only if snow is white

If we understand this claim to be about a sentence, then, assuming it is interpreted as making a necessary claim, the claim is false. There is a lot more involved than the whiteness of snow for it to be the case that 'Snow is white' is true. Tarski's insight was that the truth predicate here should be viewed as (tacitly) indexed by the interpreted language to which it belongs: it must be the case that 'Snow is white' *means* that snow is white, but this is a fact ignored by the claim. If we were to let 'Snow is white' denote a proposition – in particular, the proposition that snow is white – then the theory looks trivial since *the proposition that snow is white is defined as being true just in case snow is white*.

This triviality charge may not be particularly threatening; at least it has the advantage of being true. Furthermore, the charge of triviality may be something deflationists wish to wear as a badge of honour. There are, however, two principal reasons why deflationists fail to adopt this position:<sup>233</sup>

- (a) the triviality at issue here has its source in the concept of *proposition* and not in the concept of truth

---

<sup>233</sup> Cf. Stoltjar & Damnjanovic 2007.



(b) nothing is being said about the *theory of meaning* – an account of the connections between the sentences of natural language and the propositions expressed by them. Deflationists are engaged in a much larger philosophical project (i.e. to provide a deflationary account of *all* semantical notions) and they can only accept the triviality of their doctrine by remaining silent on an account of meaning, thereby undermining the magnitude of the program they are involved in.

Another possible response to the dilemma then would be to assume that their theory applies to *interpreted* sentences – those possessed of meaning. In this way, deflationists overcome the objection that they are ignoring the fact that sentences have meaning as it forms part of their presumptive apparatus. In this case, the deflationist would need to provide some *other* account of what it is for a sentence to mean what it does *without* employing the concept of truth, on pain of circularity. Given that most theories of meaning do in fact employ the concept of truth, this would prove to be a monumental task. And, as we have already seen, the appeal made to the use-theoretic account of meaning favoured by Horwich is fraught with all sorts of difficulties.

Unlike most deflationists, Horwich is prepared to accept that truth *is* a real and non-redundant property. The very presence of a well-defined truth predicate in our language is, in Horwich's view, sufficient to support the claim that there is an associated property of truth – though, indeed, of a peculiar kind, namely one not susceptible to “conceptual or scientific analysis”.<sup>234</sup> The minimalist idea is that all there is to the concept of truth is encapsulated by what is expressed by all instances of the equivalence schema (ES); all that a theory of truth has to offer is the (non-linguistic) proposition expressed by any such totality of instances:

...our thesis is that it is possible to explain *all* the facts involving truth on the basis of the minimal theory...the minimalist *conception*: i.e., the thesis that our theory of truth should contain nothing more than instances of the equivalence scheme.  
(Horwich 1990, p. 7-8)

This is intended to sidestep the problems of quantifying over the schema that we saw above. So, Horwich's minimal theory is the infinitely large set of (ES) instances and it is our “brute acceptance”<sup>235</sup> of (ES) instances that constitutes our understanding of truth; nothing deeper than this is involved. The word ‘true’ picks out an indefinable property of propositions and, according to Horwich, this is owing to truth's not possessing an underlying nature which philosophers are

---

<sup>234</sup> Horwich 1998a, p. 5.

<sup>235</sup> *Ibid.*, p. 13.

supposed to tease out and give expression to i.e., we *cannot* say (indeed, *should not* say) that ‘*x* is true’ is equivalent to *x* having the characteristics / property / nature of being *F*; we cannot describe the conditions under which ‘*x is true ↔ x is F*’ holds.<sup>236</sup>

Horwich also emphasizes the point that truth serves the purpose of generalization (in the minimalist case, generalization over propositions). He believes that the concept of truth provides the solution to the problem of finding “a single, finite proposition that has the intuitive logical power of the infinite conjunction”.<sup>237</sup> So, the central thrust of the minimalist thesis resides in the fact that our understanding of the truth predicate consists in our disposition to accept, without evidence, any instantiation of the equivalence schema. We do not need any more theoretical constructions in explaining truth – “the entire conceptual and theoretical role of truth may be explained on [the basis of the minimal theory]”.<sup>238</sup> Our understanding is not refined as the enumeration of the instances grow; rather our understanding consists in having the disposition alluded to. And this is part of the reason Horwich views his theory simple. In all non-trivial deployment of the truth predicate, it is acting as a device of generalization and, to do so, it is necessary and sufficient that it satisfy the equivalence schema. Hence, that schema is conceptually basic.<sup>239</sup> Its instances implicitly define truth.

We can summarize the central claims of minimalism as:<sup>240</sup>

1. Utility of the Truth Predicate: the truth predicate increases our expressive power by allowing us to express schematic generalizations by a single statement (as demonstrated above).
2. The Concept of Truth: the concept is implicitly defined by the equivalence schema and fixes the meaning of “true” (which is indefinable in that it is not amenable to reductive analyses of the type “*x is true ↔ x is F*”). We already possess a (pre-theoretic) understanding of the

---

<sup>236</sup> Compare this with our search for an understanding of what it is for something to be *red*, for instance, which would be given in terms of determinate wavelengths of light (or in terms of the set of all red things if an *extensional* definition {as opposed to an *intensional* one} was preferred).

<sup>237</sup> Horwich 1998a, p. 3.

<sup>238</sup> *Ibid.* p. 5.

<sup>239</sup> Horwich is no half-hearted minimalist! He is a minimalist to related notions such as reference and satisfaction. For instance, Horwich believes that the only viable theory of reference consists of all instances of the schema:  $\forall x (‘a’ \text{ refers to } x \leftrightarrow x=a)$ .

<sup>240</sup> Cf. Horwich 1998b (1999), p. 262. See also Lindström 2001.

truth concept and it is this grasp, which accounts for why we have a disposition to accept each, and every, instance of the equivalence scheme.<sup>241</sup> The explanatory resources that need to be called upon in explicating truth is viewed to be, therefore (and contrary to the substantialist), ‘minimal’.<sup>242</sup>

3. The Nature of Truth: truth has no underlying nature – the equivalence scheme contains *all* the basic facts about it. If we consider the following instances of the equivalence scheme:

the proposition that reality is socially constructed is true if and only if reality is socially constructed

the proposition that mathematics has no foundations is true if and only if mathematics has no foundations

the proposition that Horwich's theory of truth is not a viable theory of truth is true if and only if Horwich's theory of truth is not a viable theory of truth

then we realize that there is *nothing in common* in virtue of which these instances are made true. It seems that each is true in virtue of conditions quite independent of each other.

#### 4.3.5 *Objections*

As with the disquotational theory, the minimalist theory of truth is, *mutatis mutandis*, inconsistent. The liar paradox can quite straightforwardly be generated from (ES). The set of equivalence instances is, therefore, inconsistent. Horwich stipulates, as a refinement of his theory, that we simply reject all ‘problematic’ instances:

Given our purposes, it suffices for us to concede that certain instances of the equivalence schema are not to be included as axioms of the minimal theory, and to note that the principles governing our selection of excluded instances are, in order of priority: (a) that the minimal theory not engender ‘liar-type’ contradictions; (b) that the set of excluded instances be as small as possible; and—perhaps just as important as (b)—(c) that there be a constructive specification of the excluded instances that is as simple as possible.  
(Horwich 1998a, p. 42)

---

<sup>241</sup> “Each, and every, instance of the equivalence schema”? What part of our pre-theoretic understanding of the concept of truth is responsible for furnishing us with an acceptance of the liar-proposition instance of (ES)?

<sup>242</sup> Cf. David 1994, p. 4.

Constraint (a) is, however, ad hoc.<sup>243</sup> And even if we did throw caution to the wind and subserve “the mere desire to preserve as many instances of (ES) as possible”<sup>244</sup> then, as Van McGee has demonstrated, this “will give us too little to go on in constructing a consistent alternative to the naïve [i.e. inconsistent] theory of truth” (McGee 1992, p. 237). The idea is that there are too many ways of consistently adding instances of the *T*-scheme (or Horwich’s (ES)) to the syntax theory, most of which have consequences wholly unacceptable from an intuitive point of view, e.g. the truth of  $\lceil 2 + 2 = 5 \rceil$  is in some maximal consistent extension of Robinson *Q* arithmetic.<sup>245</sup> This is an immediate consequence of the following (Theorem 1 of McGee 1992):

Theorem 1:

Let  $\Delta$  be a set of sentences consistent with some theory *S* containing Robinson *Q* arithmetic, formulated in language *L* that contains the truth predicate  $Tr(\lceil x \rceil)$ . There is a set  $\Gamma$  of *T*-sentences such that (1) all elements of  $\Delta$  in *S* are entailed by  $\Gamma$ , (2)  $\Gamma$  is consistent with *S*, (3) every set of *T*-sentences with  $\Gamma$  as a proper subset is inconsistent with *S*, and (4)  $\Gamma \cup Q$  is a complete theory, i.e. from  $\Gamma \cup Q$ , for every sentence  $\phi$  of *L*, we can deduce either  $\phi$  or  $\neg\phi$ .

This theorem is a remarkably nice result. It rests on the following: suppose we have a first order language *L* containing the truth predicate  $Tr(\lceil x \rceil)$ , as in Theorem 1. Then, for any sentence *A* of *L*, there is a sentence *B* of *L* such that *A* is provably equivalent to  $B \leftrightarrow Tr(\lceil B \rceil)$ . This is easy to prove given Gödel’s Fixed-Point theorem, for the formula  $A \leftrightarrow Tr(\lceil x \rceil)$  is a formula of the form  $\psi(\lceil x \rceil)$  and by Gödel’s Fixed-Point theorem there is a sentence *B* which is its fixed point, i.e. a sentence *B* such that  $B \leftrightarrow \psi(\lceil B \rceil)$  is provable. Hence,  $B \leftrightarrow (A \leftrightarrow Tr(\lceil B \rceil))$  is provable, which by propositional logic is equivalent to  $A \leftrightarrow (B \leftrightarrow Tr(\lceil B \rceil))$ , i.e. *A* is provably equivalent to the *T*-schema instance involving *B*. McGee’s theorem follows from the fact that for any sentence *A* in  $\Delta$  there is a *T*-schema instance involving some *B* (as above) which entails *A*. As  $\Delta$  is by assumption consistent with *S*, then so is

---

<sup>243</sup> Indeed, one may not *even* know that the ‘liar-type’ contradictions can be engendered; the inconsistency may not be apparent.

<sup>244</sup> McGee talks of the *T*-schema (*T*). There is no logical difference between Horwich’s (ES) and the *T*-schema (*T*) – Horwich’s (ES) is (*T*) with propositions as truth bearers.

<sup>245</sup> As Gödel numbering allows us to intertranslate syntactic facts with number-theoretic ones, the task of seeking maximal sets of *T*-instances consistent with the laws of the underlying syntax can be reformulated into one seeking maximal sets of *T*-instances consistent with the basic laws of arithmetic.

the set of corresponding  $T$ -sentences equivalent to those in  $\Delta$ , and by Zorn's Lemma<sup>246</sup> this set is extendable to a maximal consistent set.

The Horwich constraint is, practically speaking, an empty one, failing to provide an acceptable deflationist theory. Imposing a 'maximalizing' constraint on the set of  $T$ -sentences gives us no guidance on how best to choose among the (uncountably) many correct (and manifestly wrong) theories. Theorem 1 holds *even if* we were to strengthen our reasoning capacities by the (superhuman)  $\omega$ -rule (this is Theorem 2 of McGee 1992):

Theorem 2:

If we replace all occurrences of "consistent" in Theorem 1 with "consistent in  $\omega$ -logic", then Theorem 1 still holds.

The addition of the  $\omega$ -rule enables one to generalize to 'All the infinite axioms of any consistent extension of Robinson  $Q$ , like  $PA$ , are true' from the fact that the truth of each is separately provable from  $PA$  together with the corresponding  $T$ -instance. These generalizations are central to the deflationist's philosophical package on *why* we have a truth predicate at all. The damage done by Theorem 2 can be seen from the fact that the application of Zorn's Lemma in Theorem 1 to generate a maximal consistent set presupposes that the underlying logic is compact, i.e. the set of  $L$ -sentences is satisfiable just in case any finite subset of  $L$ -sentences is satisfiable. If the underlying logic were not compact then one might have the union of infinitely many finite consistent subsets being inconsistent. The addition of the  $\omega$ -rule destroys compactness. And since this rule is in many ways a natural one when the syntax language is arithmetic, Theorem 1 acquires greater philosophical weight by being seen to hold even when the underlying logic is not compact.

This is, in fact, the most damaging criticism – that, like with the disquotational theory, minimalism simply fails to 'explain all the facts about truth'. The weakest chink in their armour is the fact that their theories fail to prove the very generalizations the purpose to which truth is claimed to serve. These generalizations are the deflationist's *raison d'être* for truth and the one theoretically significant role for which the deflationist acknowledges truth to be essential. Both disquotationalism and minimalism fail to prove, for example, the general fact that every true conjunction has true conjuncts. The closest they get to this is proving every instance, e.g. the instances 'if the proposition

---

<sup>246</sup> Also known as the Kuratowski-Zorn lemma, this is equivalent to the Axiom of Choice of Set Theory. It states that

every non-empty set of sets closed under unions of nonempty chains, i.e. if  $Y \subset X$  and  $Y \neq \emptyset$  and  $Y$  is a chain, then  $\bigcup Y \in X$ , has a maximal element, i.e. an element  $x \in X$  such that if  $x \subset y \in X$  then  $x = y$ .

that  $p$  and  $q$  is true then the proposition that  $p$  is true and the proposition that  $q$  is true'. The generalization and the conjunction of all its instances differ in logical strength – for one thing, the generalization entails the collection of its instances, but not vice-versa: the entailment relation is asymmetric.<sup>247</sup> As with  $DS$ ,  $(ES)$  is  $\omega$ -incomplete.<sup>248</sup> Horwich acknowledges this and proposes an escape route *via* the  $\omega$ -rule.<sup>249</sup> Horwich says, “Clearly, a set of premises attributing some property to each object of a certain kind does not entail that everything of that kind has the property”:

...for its reliability hinges not merely on the meanings of the logical constants, but also on the nature of propositions. But it is a principle we do find plausible. We commit ourselves to it, implicitly, in moving from the disposition to accept any proposition of the form ‘ $x$  is  $F$ ’ (where  $x$  is a proposition) to the conclusion ‘All propositions are  $F$ ’. So we can suppose that this rule is what sustains the explanations of the generalizations about truth with which we are concerned.  
(Horwich 1998a, p. 137)

Consider the consequences of Horwich’s proposal for adopting the  $\omega$ -rule. Recall that this rule allows the inference of the universal statement  $\forall n\Phi(n)$  from the infinitely many premises  $\Phi(0)$ ,  $\Phi(1)$ ,  $\Phi(3)$ ,... that result from replacing the numerical variable  $n$  in  $\Phi(n)$  with the numeral for each natural number. The  $\omega$ -rule is valid only if the relevant quantifier connected to it is interpreted substitutionally.<sup>250</sup> But, as already maintained, given that substitutional quantification is understood in terms of the truth of its substitution instances, the minimalist cannot help herself to this on pain of circularity. Moreover, the  $\omega$ -rule is not *even* a decidable relation between the (infinitely many) premises and conclusion of the argument Horwich gives in the first line of the quote above. It is, by nature, an *infinitary* rule of inference and, as Panu Raatikainen has remarked:

[c]onsequently, we finite human beings are never in a position to apply the  $\omega$ -rule...even if the rule would in theory entail the desired generalizations about truth, we human beings would never reach any of these generalizations. It would only be possible for an idealized infinite mathematical super-being.  
(Raatikainen 2005, p. 176)

But, clearly, we do actually reach these generalizations and so something is amiss in the deflationary proposal on truth.<sup>251</sup>

---

<sup>247</sup> This argument against deflationism, dressed in slightly different clothing, is given much weight in the next chapter.

<sup>248</sup> A theory is  $\omega$ -incomplete just in case there is a sentence  $A(x)$  with one free variable such that (i)  $A(n)$  is a theorem for every natural number  $n$ , and (ii)  $\forall xA(x)$  is not a theorem.

<sup>249</sup> Tarski made similar observations in Tarski 1936a (1983), pp. 256-258.

<sup>250</sup> See Harman 1971 and Hazen 1998.

<sup>251</sup> In accepting an argument due to Patrick Grim (Grim 1991) according to which the collection of minimalist axioms is too large to form a set, Horwich conceives his totality of minimalist axioms not as a set but as a *proper class* (Horwich

This chapter has outlined a number of weaknesses in the deflationary thesis about truth. There are quite a few compelling reasons already to reject this thesis in its various forms and lay it to rest. The purpose of the following chapter is to hammer the final nail in the deflationist's coffin.

---

1998a, p. 20). The intended universe of discourse is, therefore, uncountable. But, how is this supposed to square with the  $\omega$ -rule – a rule requiring that for every element in the intended domain there is a canonical name? Grim's argument is similar to Cantor's paradox concerning 'the' totality of all sets and the Russellian conclusion for the *non-existence* of such a set. Lindström 2001, pp. 174-177, cites other question-begging issues on the similarity in approach adopted by Russell and that adopted by Horwich. Crucially, Russell and (Grim) direct their attention to the *well-definedness* of the totality of propositions that it makes sense to quantify over, and not against the assumption that such a totality forms a set – something that Horwich assumes in applying Cantor's argument in conceiving of his minimalist totality as a proper class.

# CHAPTER 5

## DEFLATIONISM AND CONSERVATIVENESS

If truth is admitted into the induction principle, it is counterintuitive to maintain that the notion is thin and insubstantial. How thin can the notion of arithmetic truth be, if by invoking it we can learn more about the natural numbers?

**Stewart Shapiro, *Proof and Truth: Through Thick and Thin***

As the discussion in Chapter 4 has shown, deflationism about truth is neither a single philosophical standpoint nor a uniform position. Rather, it is an assortment of disparate, and mutually incompatible, doctrines. Deflationism is a broad church and best characterised as more of a tendency than anything else. Different deflationists formulate this tendency variously, placing emphases differently, thus making it difficult to garner what may be considered as unshakably essential and what optional. As Crispin Wright suggests, this makes it “...a rather more complicated business to elicit what is fundamentally unsatisfactory about the deflationary conception of truth”.<sup>252</sup>

Nevertheless, recently, Stewart Shapiro and Jeffrey Ketland have proposed a set of common principles. They maintain that, given the deflationary tendency, the picture deflationists paint comfortably fits into a *single* semantical picture. If what is at the heart of any deflationary view of truth is the claim that the concept of truth is not essential to explanations, that it is dispensable if one argues for a substantial claim because it serves (like the logical connectives) a logico-linguistic function only, then this amounts to the claim of the *conservativeness* of the axioms and rules for truth. This chapter demonstrates that the deflationist would be hard-pressed to deny such a commitment and that this ultimately proves to be her undoing.

### 5.1 *The Conservativeness Implication*

Ketland 1999 sees an analogy with deflationism about mathematics – the *dispensability* programme initiated by Hartry Field in 1980 with his *Science Without Numbers: A Defence of Nominalism*. This is a response to a challenge laid down by Hilary Putnam in his 1971 thesis *Philosophy of Logic*<sup>253</sup> to translate Newtonian Gravitational Physics into ‘nominalistic language’.<sup>254</sup> A central component of

---

<sup>252</sup> Wright 1999, p. 209.

<sup>253</sup> Reprinted in Putnam 1975b (1979)).

<sup>254</sup> A language in which talk of abstract universals is to be replaced by talk of linguistic entities, namely predicates.



Field's work is to show that adding mathematical axioms  $M$  to a 'mathematics-free' or 'nominalistic' theory  $N$  would yield a *conservative extension* of  $N$ , i.e.:

...any inference from nominalistic premises to a nominalistic conclusion that can be made with the help of mathematics could be made (usually more long-windedly) without it.<sup>255</sup>  
(Field 1980, p. x)

It follows that the addition of the mathematics  $M$  is redundant in explaining or deducing the phenomenon described by any logical consequence of  $N$ .<sup>256</sup> As Ketland 1999 points out, a deflationist about truth might put her view similarly:

Any inference from non-semantical premises to a non-semantical conclusion that can be made with the help of a truth theory can be made without it.

In particular, a theory of truth is conservative if it does not prove new theorems that do not contain the truth predicate and which are not provable without the axioms of the truth theory – adding the truth predicate, with rules essential to it from the deflationary point of view, should not imply *anything* not related to truth. It is difficult to see how this conservativeness thesis could *not* be a consequence of the deflationary tendency. As Shapiro remarks:

...it seems that in some sense or other, the deflationist is committed to the conservativeness of truth. Deflationism presupposes that there is some sense of 'consequence' according to which truth is conservative.  
(Shapiro 1998, p. 498)

Conservatism is just what we ought to expect from a position opposed to 'substantial' theories of truth. An *insubstantial* theory ought to only have insubstantial consequences, not contributing any substantial insights to what we already know, i.e. the theory should not imply any new non-

---

<sup>255</sup> So, if we let  $N$  be a 'mathematics-free' theory of the natural world,  $M$  be a standard mathematical theory and  $\phi$  be any mathematics-free assertion. Then, if  $\phi$  is a logical consequence of  $N \cup M$ ,  $\phi$  is a logical consequence of  $N$ .

<sup>256</sup> It could be argued (and indeed Field 1980 has this suggestion) that what 'deflationary' mathematicians require is that the addition of mathematical axioms is dispensable for *theorem proving*, i.e. if  $\phi$  is a theorem of  $N \cup M$  then  $\phi$  is a theorem of  $N$ . But Shapiro 1983 shows that this metatheorem is not true. This deals with whether the base theory (to which the mathematical axioms are being 'added') has a language containing *axiom schemes*. If we include formulas containing vocabulary from the extended language (i.e. from the language of the combined theories) in forming instances of the axiom schemes of  $N$  (and the schemes are treated as *lists*) then the addition of  $M$  to  $N$  yields a conservative extension. If the schemes are treated as *rules*, however, then the addition yields a non-conservative extension. Shapiro illustrates this in the context of Field's effort to nominalize the theory of gravitation in flat Euclidean space-time. Ketland 1999 deals with this subtlety.

semantic consequences. If we start off with a base theory  $\Gamma$  and add to it its theory of truth i.e. all  $T$ -sentences  $Tr(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$ , for sentences  $\varphi$  in the language of  $\Gamma$ , *and nothing else*, it is essential for the deflationist that this extension be conservative over the base theory  $\Gamma$ ; otherwise there would be statements in  $\Gamma$  that could not have been established without the help of the extension. If that were to be the case, then the deflationist would be forced to admit that the truth predicate has added content to our base theory, and is thus more committal than it was previously understood to be.

Michael Sheard first raised the issue of conservativity, but without making any (explicit) connection to the insubstantialist-substantialist philosophical debate:

...does the fact that a particular theory is a conservative extension of the base theory in a language without the truth predicate indicate that we have selected the ‘appropriate’ additional axioms for truth itself, or should we expect the introduction of a theory of truth to provide more formal power with respect to the underlying domain?  
(Sheard 1994, p. 1053)

and perhaps it is owing to this philosophical connection not being made explicit that many working in this area claim that they are “not aware of any deflationist who has explicitly bound him- or herself to the doctrine that her theory of truth is conservative”.<sup>257</sup> But, as Shapiro puts it:

I do not know if there is a consensus among deflationists over how metaphysical substance lines up with expressive and proof-theoretic power, but there is at least cause for thought here.<sup>258</sup>  
(Shapiro 1998, p. 495)

Besides, as Ketland notes, despite this non-committal attitude the (restricted)  $T$ -scheme’s conservativeness *does* furnish the deflationist with a justification for the kind of metaphysics she wishes to advance:

...the fact that the (restricted)  $T$ -scheme *is* conservative provides an important analysis of a number of features of the truth predicate that the deflationist wants, such as:

1. The *dispensability* of the truth predicate;
2. the *epistemological neutrality* of truth;
3. the *contentlessness* of truth.

...this is *exactly* the sort of (formal) behaviour that the deflationist wants (or should want) from a truth theory.  
(Ketland 2000, p. 320)

---

<sup>257</sup> Halbach 2001a, p. 167.

<sup>258</sup> Shapiro continues: “Why is it that the metaphysically thin, or natureless, or lightweight concept of truth should be a sure-fire sign of expressive and proof-theoretic strength?”

There is a further bonus in deflationary truth being logically conservative: it is demonstrably consistent.<sup>259</sup> Unless the base theory already happens to be inconsistent, no contradictions can be generated from a conservative truth theory. Hence, the view that deflationism about truth entails conservativity ought to be especially attractive to those wishing to advance a lean concept of truth.

But deflationism is a stronger claim than mere subscription to conservativity. According to Paul Horwich, the totality of all instances of the equivalence scheme yield all there is to say about the concept of truth. For example, Horwich writes:

...our thesis is that it is possible to explain all the facts involving truth on the basis of the minimal theory...i.e., the thesis that our theory of truth should contain nothing more than instances of the equivalence scheme.  
(Horwich 1990, pp. 7-8)

Another deflationist, Michael Williams, agrees:

[Deflationists] think that when we have pointed to certain formal features of the truth-predicate (notably its 'disquotational' feature) and explained why it is useful to have a predicate like this (e.g. as a device for asserting infinite conjunctions), we have said just about everything there is to be said about truth.  
(Williams 1988, p. 424)

We can sum this up in the following:

The Completeness Condition: There is nothing more to the concept of truth than is given by the theory consisting of all instances of the *T*-schema, deflationarily understood. Such a theory is a *complete* theory of the concept of truth.

The remainder of this chapter will be devoted to demonstrating that the deflationary thesis, so understood, *is false*. Before doing so it will be useful to link the discussion to that famous foundationalist movement, Hilbert's Programme.

---

<sup>259</sup> A point that David Hilbert understood in his attempt to establish mathematics on a secure footing with his Consistency Programme - see below.

### 5.1.1 *An Analogy with Hilbert's Programme*

The idea that certain philosophical claims can be rendered harmless by making use of conservativeness arguments dates back to the turn of the century, to David Hilbert. Hilbert was concerned to show (among other things) that the use of certain abstract (*ideal*) entities in mathematics is legitimate. The intellectual climate of the time was one in which mathematicians were preoccupied by foundational concerns. One in particular was the problem of the legitimacy of abstract objects.<sup>260</sup> Despite the work done by Weierstrass in clarifying the role of the infinite in analysis and Cantor's systematic theory of transfinite numbers, many influential figures in the mathematical community (in particular Kronecker, Poincaré and Brouwer) challenged the validity of infinitistic reasoning. Hilbert, however, was determined to save what he called 'Cantor's Paradise'. Speaking of Brouwer, Weyl (Brouwer's protégé) and Kronecker, Hilbert said:

They seek to save mathematics by throwing overboard all that which is troublesome... They would chop up and mangle the science. If we would follow such a reform as the one they suggest, we would run the risk of losing a great part of our most valuable treasures!  
(Reid 1970, p. 155)

A recent commentator explains Hilbert's view in even more dramatic (even melodramatic) language:

Actually, Hilbert saw the issue as having supramathematical significance. Mathematics is not only the most logical and rigorous of the sciences but also the most spectacular example of the power of "unaided" human reason. If mathematics fails, then so does the human spirit. I was deeply moved by the following passage: "The definitive clarification of the nature of the infinite has become necessary, not merely for the special interests of the individual sciences but for the honour of human understanding itself."  
(Maddy 1988, p. 492)

The Intuitionists, by contrast, possessed an extremely revisionary attitude. Hilbert, however, wanted to rescue classical mathematics. He argued that none of mathematics – and not just the finitistic part – is to be rejected. Hilbert sought to justify infinitistic mathematics, and attempted to demonstrate that finitistic mathematics can be used to this end. Hilbert realized that if he could show by finitistic means that the use of abstract techniques is conservative then one need not worry about such abstractions. We might call this Hilbert's 'Conservation Programme'. More precisely, all he needed to concern himself with were the finitistically meaningful statements of concrete mathematics (the *real* statements involving reasonably simple functions). The abstract statements involving 'ideal'

---

<sup>260</sup> The fires of this controversy were fanned by the revolutionary developments taking place simultaneously in the world of physics.

entities were to have no independent meaning but *were*, nonetheless, candidates for abstract manipulation.<sup>261</sup> The ideal statements and the abstract reasoning would have to be, according to Hilbert, codified in some formal system with effective rules for the combinatorial manipulation of symbols. The reason for this is that since proving the *consistency* of the formal system encoding the abstract concepts involves simple combinatorial manipulations, this already helps establish the sought conservation result.<sup>262</sup>

Sadly, in 1931, Kurt Gödel proved a couple of incompleteness theorems delivering the *coup de grâce* to Hilbert's Consistency Programme – and, by implication, his Conservation Programme. In particular, according to Gödel's second incompleteness result, if  $T$  is a consistent, sufficiently rich formal theory, then  $T$  cannot prove a sentence asserting its own consistency.<sup>263</sup>

## 5.2 Is Deflationary Truth Conservative?

In setting up formal models of truth with respect to a formal system  $S$ , it is often taken to be a predicate applying to the Gödel numbers, or other appropriate codings, of sentences which allow those sentences, and their properties, to be discussed in  $S$ . Peano arithmetic (henceforth  $PA$ ) is commonly chosen as the theory of those objects to which truth applies principally because it is equivalent to simple theories of syntax that leave the structure of the underlying syntax bare and perspicuous. But the choice of  $PA$  is not inevitable:<sup>264</sup> our investigation could equally be conducted with set theory as the base theory. However, arithmetic is chosen because it is the most formally economical system,<sup>265</sup> and has moreover proved versatile in yielding many interesting systems once the truth-theoretic axioms have been added.  $PA$  will serve as the base theory in this section.<sup>266</sup>

---

<sup>261</sup> Such as the imaginary number  $i$ . Though not a real number,  $i$  can be manipulated algebraically ( $i^2 = -1$ ) and its use leads to no new algebraic identities among real numbers.

<sup>262</sup> See the appendix to this chapter for a demonstration of how this works.

<sup>263</sup> There are infinitely many such sentences: see e.g. Mostowski 1966, Lecture II.

<sup>264</sup> Although, as Shapiro notes, “any serviceable theory of strings will have analogues of the basic Peano axioms”. (2002, p. 110). For example, this is the analogue of the induction scheme in string theory: if  $\Phi$  holds of the null string, and if, for any string  $z$ , if  $\Phi(z)$  then  $\Phi$  holds of the result of adding a character to  $z$ , then  $\Phi$  holds of all strings.

<sup>265</sup> Although this does make it correspondingly farther removed from intuition, which is why Tarski himself used set theory.

<sup>266</sup> There is a discussion in the literature on whether deflationary truth entails conservativeness over *logic* or over the underlying syntax. According to Halbach, “[C]onservativeness over logic guarantees the complete neutrality of the truth theory with respect to any non-semantic questions, while conservativeness over a theory of expressions... renders the axioms neutral with respect to questions that are left undecided by the theory of expressions. Therefore, conservativeness over logic is more desirable.” (2001a, p. 168). However, a trivial argument serves to establish that even the most modest theory of truth (the restricted set of  $T$ -sentences: ‘ $A$ ’ is true iff  $A$ , where  $A$  is any sentence not

As far as the (restricted) set of  $T$ -sentences is concerned, adding them to  $PA$  is indeed a conservative extension;<sup>267</sup> i.e. no more arithmetical sentences are derivable than those already derivable in  $PA$ .<sup>268</sup>

If we let  $PA(T)$  represent the system  $PA +$  (restricted)  $T$ -scheme then the following fact obtains:

Fact 1:- The system  $PA(T)$  is a conservative extension of  $PA$ .<sup>269</sup> This means

---

containing the truth predicate) is *not* conservative over logic. If we let  $\Psi$  be any tautological sentence of the base language with  $n$  a name (or code) of  $\Psi$  and  $m$  a name (or code) of  $\neg\Psi$ , then the following are the corresponding  $T$ -sentences associated with  $\Psi$  and  $\neg\Psi$ :

$$\begin{aligned} Tr(n) &\leftrightarrow \Psi \\ Tr(m) &\leftrightarrow \neg\Psi \end{aligned}$$

Since  $\Psi$  is provable in first-order logic with identity, and  $\neg\Psi$  refutable in it,  $Tr(n)$  and  $\neg Tr(m)$  are logical consequences of our modest theory of truth. This means that  $n \neq m$  is also a logical consequence. So, our modest theory has the following claim as a consequence:

$$\exists x \exists y (x \neq y)$$

which is not a logical truth. Any adequate theory of truth entails that there are at least two objects in the universe. (thought in his corrigendum to Ketland 1999, Ketland does provide a proof according to which the addition of the  $T$ -sentences to any theory entailing  $\exists x \exists y (x \neq y)$  is a conservative extension of that theory). Hence, as Halbach puts it, “*the deflationist has to accept that even the most mundane axioms for truth have ontological impact*” (*Ibid.*) This means that the requirement that a correct theory of truth be conservative over logic puts too stringent a demand on the deflationist (*cf.* Shapiro 1998, pp. 497-498). Unless ontological consequences are considered an intrinsic feature of logic, the concept of truth cannot be understood as a logical constant as the deflationist intends it to be. If such an impact is permissible, then it is difficult to know what logic is. Truth, therefore, is unlike the logical connectives and quantifiers that serve their logical purpose in an ontologically neutral fashion.

Given that in formulating the above  $T$ -sentences both were assumed to have names (or codes in the envisioned Gödel-numbering), that the theory logically implies the existence of at least two different objects should come as no surprise. However the formula  $\Psi$  (our truth-bearer) is to be understood – proposition, utterance, sentence type, etc. – the background syntax ought to entail that  $\Psi$  is a *different* sentence from  $\neg\Psi$ . In other words: “[A] *truth theory makes sense only if it comes with a theory of the objects that may be true, whether they are sentences, propositions, or something else. Truth theoretic axioms should not be judged when separated from their underlying ontological theories that serve as base theories... That certain presuppositions (made in motivating the axioms for truth) drop out again from these axioms should hardly be surprising. This means that [deflationists] asked for the wrong kind of conservativeness before. If truth is not substantial, then it should not imply anything beyond the presuppositions that have been made when setting it up. The truth theory should not produce more than one has sunk into it.*” (Halbach 2001a, pp. 181-182).

<sup>267</sup> Ketland 1999, Theorem 1. The model-theoretic proof is sketched as follows: Any model  $M$  of  $PA$  can be expanded to a model  $(M, E)$  which satisfies the restricted  $T$ -scheme by taking the interpretation of the truth predicate  $Tr$  (our  $E$ ) to be the set of (codes of)  $\mathcal{L}_{PA}$ -sentences true in  $M$ . This in turn implies that if we can deduce an arithmetical sentence  $\phi$  from  $PA(T)$ , then we can deduce  $\phi$  from  $PA$  itself.

<sup>268</sup> Moreover, even if we were to choose only instances of the induction scheme in  $PA$  as the axioms of our base theory and extend the language to include the  $T$ -sentences together with truth-theoretic instances of the scheme (i.e. not just the arithmetical ones), conservativity is maintained. We can go further: “*One can...strengthen this truth theory without losing conservativeness by replacing the  $T$ -sentences with their uniform counterparts: the uniform  $T$ -sentences saying that for any  $x$ ,  $A(x)$  is true for  $x$  if and only if  $A(x)$ .*” Halbach 1999b has the proof. However, note that not every model of  $PA$  can be expanded to a model of  $PA +$  truth-theoretic induction axioms + uniform  $T$ -sentences. (Kaye 1991, Proposition 15.4).

<sup>269</sup> In fact, as already mentioned, extending the language of  $PA$  to include the non-truth-theoretic instances of the  $T$ -scheme, means that the resulting theory  $PA(T)$  is consistent (assuming the consistency of  $PA$ ). Note that, according to Paul Horwich, we should not restrict the  $T$ -scheme apart from when pathology is imminent i.e., iterations of the truth predicate *are* (sometimes) permissible, though sometimes not.

that, if  $\phi$  is an arithmetical sentence and  $PA(T) \vdash \phi$ , then  $PA \vdash \phi$

Even if we were to allow induction axioms that incorporate semantic reasoning by permitting the inclusion of the truth predicate, then  $PA(T)$  generates a conservative extension of  $PA$ . Do these results mean that the deflationist can rest content? Can she vindicate her position by alluding to these conservativeness facts? Unfortunately, no she cannot. This is owing to the deflationist's commitment to a theory of truth much stronger than that recommended by the restricted set of instances of  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ . Trouble for the deflationist begins when it is recognized that truth-theoretic equivalents of infinite conjunctions are *not* derivable from the theory consisting of the  $T$ -sentences. A central deflationary contention is that part of the logical *raison d'être* of the truth predicate is to endorse a potential infinite list of assertions like 'All logical consequences of true statements are true' or 'All theorems of  $PA$  are true'.<sup>270</sup> Generalizations of this sort are precisely the theoretical purpose the deflationist says is discharged by truth:

Fact 2:- The system  $PA(T)$  does not imply 'All theorems of  $PA$  are true'

Moreover, there are certain generalizations about truth that one would like to establish from this proposed theory of truth which (provably) cannot be established. For example, this theory of truth cannot prove the truth-theoretic law of excluded middle nor prove that a sentence is not true if its negation is true.

One natural proposal for strengthening the (restricted) set of  $T$ -sentences is to pick the Tarskian inductive definition of truth for arithmetic. This takes the form of six inductive axioms:

( $T_{At}$ ) If  $\phi$  has the form  $t = u$ , then  $\phi$  is true iff  $val(t) = val(u)$

( $T_{\neg}$ )  $\neg\phi$  is true iff  $\phi$  is not true

( $T_{\wedge}$ )  $\phi \wedge \psi$  is true iff both  $\phi$  and  $\psi$  are true

( $T_{\vee}$ )  $\phi \vee \psi$  is true iff either  $\phi$  is true or  $\psi$  is true

( $T_{\forall}$ )  $\forall x \phi$  is true iff, for each number  $n$ ,  $\phi(\tilde{n})$  is true

( $T_{\exists}$ )  $\exists x \phi$  is true iff, for some number  $n$ ,  $\phi(\tilde{n})$  is true

[where the  $\phi$ 's and  $\psi$ 's are arithmetical sentences only]<sup>271</sup>

<sup>270</sup> Halbach shows how to produce truth-theoretic equivalents of infinite conjunctions (1999a, Proposition 1).

<sup>271</sup> The axiom '( $T_{Sent}$ ) If  $\phi$  is true, then  $\phi$  is a sentence of  $\mathcal{L}_{PA}$ ' is sometimes added.

By induction on formula complexity, the theory given by the axioms of  $PA$  + arithmetical induction axioms + the above axioms governing the truth predicate implies each (arithmetic) instance of the  $T$ -scheme. So, this meets Tarski material adequacy condition. Let us call this theory  $T(PA)$ . The first important result to note is the following:

Fact 3:-  $T(PA)$  is a conservative extension of  $PA$ <sup>272</sup>

These inductive clauses for truth ought to be deflationarily acceptable. Indeed, Field embraces them, arguing that “it is clear that without such general laws the truth predicate would not serve its main purpose [of allowing us to make natural generalizations we could not otherwise make]”.<sup>273</sup> However, a question mark still lingers. An appeal to Tarski’s theory here already brings with it stronger ideological commitments. Moreover, it is stronger than the pure (restricted) disquotational  $T$ -sentence theory as, using just the  $T$ -sentences, you cannot prove things like “For any sentence  $x$  of  $\mathcal{L}_{PA}$ , the negation of  $x$  is true iff  $x$  is not true”. There is a case, then, to be made that appealing to Tarski’s inductive definition moves the deflationist into strictly *non-deflationary* territory. The deflationist’s position is forced into untenability here. The conservativeness of  $T(PA)$  ensures that certain natural generalizations, like the soundness principle for  $PA$  are unprovable in the extended theory:

Fact 4:- As a corollary of Fact 3,  $T(PA)$  does not imply ‘All theorems of  $PA$  are true’<sup>274</sup>

Furthermore, once the truth predicate is added to our extended language, with quantificational assistance and by using the provability predicate that  $PA$  possesses, we can express generalizations

---

<sup>272</sup> Kotlarski, H., Krajewski, S. & Lachlan, A. 1981 have shown how to establish this result model-theoretically. The idea is to expand a *countably* recursively saturated model of  $PA$  to a model of  $T(PA)$ , to overcome the problem that only recursively saturated models can be expanded. This method works because every countable model of  $PA$  has an elementarily equivalent extension that is recursively saturated. Halbach 1999b, however, establishes this result by cut elimination, an argument that can be carried out in  $PA$  itself, thus allowing a consistency proof of  $T(PA)$ , under the assumption that  $PA$  is itself consistent.

<sup>273</sup> Field 1998, p. 535.

<sup>274</sup> Moreover, the reflective proposition that ‘All theorems of  $PA$  are true’ is provably stronger than  $PA$ . So, the deflationary claim that an acceptance of  $PA$  and its corresponding truth-claim are logically equivalent is mistaken. For instance, from the reflective proposition you can prove (under the assumption that  $PA$  is consistent) the consistency statement of  $PA$ , together with  $PA$ ’s Gödel sentence,  $G_{PA}$ . And, you cannot do this with  $PA$  alone.



like ‘All axioms of  $PA$  are true’<sup>275</sup> or ‘All the rules of inference of  $PA$  preserve truth’. An adequate theory of truth (deflationary or otherwise) should have the resources to *establish* such claims. The deflationist is met with difficulty in that such generalizations are not capable of being established in this proposed extended theory. And recall that the provision of such logico-linguistic fertility is precisely the *raison d’être* conferred on truth by the deflationist.

The deflationist may wish to avail herself of the following fact. By including the truth predicate in the induction scheme of  $PA$ , she can, with the help of this enlarged system, prove the requisite generalizations. The problem, however, is that this enlarged theory is *not* a conservative extension of  $PA$ . Let us call the theory of  $PA$  (with the induction scheme expanded to admit formulas containing the truth predicate) + the Tarskian ‘inductive’ clauses for truth,  $Tr(PA)$ .<sup>276</sup> We have the following:

Fact 5:-  $Tr(PA)$  is not a conservative extension of  $PA$

Fact 6:-  $Tr(PA)$  implies soundness/reflection principles like ‘All theorems of  $PA$  are true’

Fact 7:- From ‘All theorems of  $PA$  are true’ you can prove (assuming  $PA$  is consistent) the consistency statement  $Con(PA)$  for  $PA$

The presence of the truth predicate in the first order induction scheme of  $PA$  endows the reasoner with the ability to generate reflective results. Moreover, the consistency statement  $Con(PA)$  expressible in the language of  $PA$  is, by Gödel’s second incompleteness theorem, unprovable in  $PA$ . But, from ‘All theorems of  $PA$  are true’, we can prove  $Con(PA)$  (if  $PA$  is consistent). If we can prove ‘All theorems of  $PA$  are true’ then we must have generated a *non-conservative* extension. Hence,

(a) Acceptance of the axioms of  $PA$ , and

(b) Accepting the soundness claim that ‘All theorems of  $PA$  are true’,

---

<sup>275</sup> There is a subtlety here that there are many formulas  $Pr(x,y)$  which represent the relation ‘ $x$  is the code of a proof of the formula with Gödel number  $y$ ’ and they are not all equivalent. One has to distinguish between these since not all express this relation in any natural sense.

<sup>276</sup>  $Tr(PA)$  is Feferman’s (1991) terminology. Ketland 1999 and Kaye 1991 call this theory  $PA(S)$ .  $Tr(PA)$  and  $PA(S)$  are intertranslatable.

which are considered to be equivalent in logical strength by the deflationist, are provably not so. (b) is a logically much stronger claim. As Ketland 1999 shows, Tarski's inductive definition of truth is adequate in this sense (i.e. non-conservative). However, this is crucially in contradiction with the deflationary requirement that her axioms and rules of truth conservatively extend the theory subscribed to. Indeed, this helps to lend further credence to the Tarskian claim that his semantic conception of truth is doing justice to the 'classical' correspondence theory of truth. And so, it would appear that substantial assumptions have been smuggled in through the back door.

### 5.2.1 Tennant's Rebuttal

The topic of whether one needs to license reflective reasoning truth-theoretically has recently been vociferously discussed in the *Mind* journal of philosophy. As explained above, Ketland corners the deflationist into occupying one of the following two positions:

- (i) abandon the conservation condition and commit to a substantialist construal of truth, or
- (ii) abandon the adequacy condition, offering a non-truth-theoretic understanding of the fact that one's acceptance of a mathematical base theory such as *PA* commits one to an acceptance of a number of further statements in the language of that theory (one of these of course being the Gödel sentence  $G_{PA}$ )<sup>277</sup>

Neil Tennant (2002, 2005) believes that option (ii) is available to the deflationist. He claims that the deflationist has properly deflationary means for attaining the insight that the undecidable Gödel sentence  $G_{PA}$  is one we ought to accept (*cf.* Tennant 2002, p. 553). One suspects Tennant believes further that there are deflationarily licit means by which to prove the soundness claim, though this is nowhere explicitly mentioned in his *Mind* responses to Ketland, for that would indeed constitute his occupying a counter-position to Ketland's. If Tennant is right, and he does indeed legitimately occupy such a counter-position, then clearly the Ketland-Shapiro argument fails. Let us closely examine what Tennant has to say on behalf of the deflationist.<sup>278</sup>

---

<sup>277</sup> Ketland calls this commitment a "*conditional epistemic obligation*" (Ketland 2005, p. 80) and, according to him, this is to be explained using the notion of truth. And in so doing, allies the concept of truth with its non-conservativeness, i.e. substantiality.

<sup>278</sup> Tennant is actually a substantialist about truth. He says: "*I am not a deflationist. I believe that truth and falsity are substantial.*" However, Tennant believes that there is still a role for a devil's advocate to play: "*The devil – an unwitting client, for whom I [am] appearing pro bono – [is] the deflationist.*" (Tennant 2005, p. 89).

In “Deflationism and the Gödel Phenomena” Tennant is troubled by what he refers to as ‘the substantialist dogma’ pervading the philosophical community and according to which only a substantial notion of truth permits one’s being able to establish the truth of the Gödel sentence  $G_{PA}$ . Tennant cites Michael Dummett as an example of a philosopher falling foul of such dogmatism, citing the following passage from Dummett’s “The Philosophical Significance of Gödel’s Theorem” as evidence:

The statement  $[G]$  is of the form  $\forall xA(x)$ , where each one of the statements  $A(0), A(1), A(2), \dots$  is true: since  $A(x)$  is recursive, the notion of truth for these statements is unproblematic. Since each of the statements  $A(0), A(1), A(2), \dots$  is true in every model of the formal system, any model of the system in which  $[G]$  is false must be a non-standard model... whenever, for some predicate  $B(x)$ , we can recognize all of the statements  $B(0), B(1), B(2), \dots$  as true in the standard model, then we can recognize that  $\forall xA(x)$  is true in that model. This fact... we know on the strength of our clear intuitive, conception of the structure of the model.  
(Dummett 1963 (1978), p. 191)

Tennant refers to this as the ‘semantical argument’ and it is clearly designed to help one understand why an acceptance of the Gödel sentence  $G_{PA}$  is the right thing to do. On behalf of the deflationist, Tennant seeks to arrive at this very understanding by avoiding any use of the truth predicate and to thereby demonstrate that this insight is deflationarily attainable (*cf.* Tennant 2002, p. 562). However, immediately alarms bells should be ringing. The point that Ketland and Shapiro make is that the reflective proposition ‘All theorems of  $PA$  are true’ and acceptance of  $PA$  are not equivalent; there are commitments that logically follow from the reflective proposition that do not follow from a mere acceptance of  $PA$ . The reflective proposition is *deductively stronger*, contrary to what the deflationist contends. There is a *logical difference* in accepting  $PA$  and accepting  $PA$ ’s soundness statement. So, to argue that  $G_{PA}$  can follow from the soundness of  $PA$  in a deflationarily licit fashion is not relevant to the dialectic here. What is important is the manner by which we arrive at the soundness claim. And this is significantly relevant to the dialectic because the deflationist claims that the  $T$ -sentences *entail all there is to be said about truth, and that all the facts about truth can be explained solely on the basis of them*. But, as demonstrated above, this is only licensable with a ‘thick’, robust, substantial, non-conservative notion of truth. In fact, Tarski had already pinned down the result that reflection principles for our non-semantical base theory  $PA$  are provable from his theory (which is non-conservative) via a proof of the soundness principle. In his 1936a *locus classicus*, he says:

The definition of truth allows the consistency of a deductive science to be proved on the basis of a metatheory which is of higher order than the theory itself. On the other hand, it follows from Gödel’s investigations that it is in general impossible to prove the consistency of a theory if the proof is sought on the basis of a metatheory of equal or lower order. Moreover, Gödel has given a method for constructing

sentences which – assuming the theory concerned to be consistent – cannot be decided in any direction in this theory. All sentences constructed according to Gödel’s method possess the property that it can be established whether they are true or false on the basis of the metatheory of higher order having a correct definition of truth. Consequently, it is possible to reach a decision regarding these sentences, i.e. they can be either proved or disproved.

(Tarski 1936a (1983), p. 274)

Nevertheless, Tennant still persists in wishing to demonstrate that there exist deflationarily licit means by which to carry out the semantical argument for the truth of the Gödel sentence  $G_{PA}$ .

Tennant adopts what he calls ‘the principle of uniform primitive recursive reflection’:

$$(UR_{p.r.}): \forall x(Prov_{PA}(\ulcorner \phi(\text{dot}(x)) \urcorner) \rightarrow \forall x\phi(x) \text{ [with } \phi(x) \text{ a primitive recursive predicate]}^{279}$$

The union of our non-semantical, mathematical base theory  $S$  and  $(UR_{p.r.})$  does indeed prove the Gödel sentence  $G_{PA}$ , i.e.  $PA \cup (UR_{p.r.}) \vdash G_{PA}$ .<sup>280</sup> And clearly allusions to alethic concepts are not involved in this reflection principle.<sup>281</sup> According to Tennant, *pace* Ketland, this furnishes us with another way to recognize the truth of Gödel sentences without invoking anything like Tarski’s theory (*cf.* Tennant 2002, p. 567). Tennant is clearly concerned by Ketland’s claim that “our ability to recognize the truth of Gödel sentences involves a theory of truth (Tarski’s) which significantly transcends the deflationary theories” (Ketland 1999, p.88). Strangely, it is to *this* claim, on behalf of the deflationist, that Tennant addresses himself.<sup>282</sup> However, as has already been highlighted, this is completely beside the point! Ketland’s (and Shapiro’s) aim is to investigate whether the justificatory resources needed to prove truth-theoretic generalizations requires, as the deflationist contends, nothing more than the  $T$ -sentences. Proving that something like Tarski’s compositional theory of truth meets the adequacy condition also shows that the conservation condition is violated and, hence, that deflationism about truth is mistaken. Moreover, the meeting of the adequacy condition yielded by employing Tarski’s theory helps explain/justify the ‘reflective reasoning’ that

<sup>279</sup> This is Smoryński’s “First Uniform Reflection Principle”. See Smoryński 1977, p. 845. It originates with Solomon Feferman in Feferman 1962. The notation  $\ulcorner \phi(\text{dot}(x)) \urcorner$  indicates that the variable  $x$  must range over elements namable by constants.

<sup>280</sup> Ketland is completely correct to point out here: “*It is acceptable to finitists, intuitionists, deflationists, nihilists, deconstructivists, etc. that  $G_{PA}$  follows from the soundness of  $PA$ .... [b]ut what does this have to do with what Shapiro and I discuss?*” (Ketland 2005, p. 82).

<sup>281</sup> Note significantly that contrary to some commentaries Tennant is *not* wishing to point out, on behalf of the deflationist, that the truth of the Gödel sentence  $G_{PA}$  follows from the conjunction of  $PA$  with the statement that  $PA$  is consistent, i.e.  $PA + Con(PA)$ . To suggest otherwise would be to misunderstand Tennant. In fact, Tennant himself flatly contradicts commentaries that do wish to point out, on his behalf, that the truth of the Gödel sentence  $G_{PA}$  follows from  $PA + Con(PA)$ : “*We are requiring that our regimentation...of the semantical argument be faithful to its informal structure. This requirement also rules out making any use of  $Con(PA)$ .*” (Tennant 2002, p. 563. Emphasis added.).

<sup>282</sup> Tennant repeats his being concerned by this alone in 2005 paper, p. 93.

begins with an acceptance of  $PA$  and concludes with an acceptance of the global reflection principle.<sup>283</sup> According to Shapiro and Ketland, this is due to an understanding of the notion of truth for the base language. By contrast, Tennant does not proffer any justification for his principle of uniform primitive recursive reflection,  $(UR_{p.r.})$ . He just posits it.<sup>284</sup> No motivation is provided for it in his 2002 paper. More significantly, he does not show that such a justification *need not be* truth-theoretic. This is what an advocate on behalf of the deflationist ought to be concerned by in order to meet Ketland's anti-deflationist challenge.

In his 2005 reply to Ketland, in highlighting that the adoption of all instances of his principle of uniform recursive reflection with the non-semantic base theory  $PA$  affords a proof of the Gödel sentence  $G_{PA}$ , Tennant maintains the following:

There is no truck here with a substantial notion of truth. All one is doing here is applying the reflective thought involved in the soundness principle...[t]here is only reflection on one's present axiomatic and deductive commitments...[n]o further justification is needed for the new commitment made by expressing one's earlier commitments. As soon as one appreciates the process of reflection, and how its outcome is expressed by the reflection principle, one already has an explanation of why someone who accepts  $PA$  should also accept all instances of the reflection principle...[w]e do not need a substantial concept of truth in order to establish this result...  
(Tennant 2005, p. 92)

Elsewhere, Tennant writes that one alternative to justifying reflection principles non-truth-theoretically is provided by “engaging in suitable intellectual reflection...[so] one can regard the theorizing about truth as a ladder to be kicked away by the deflationist who has discovered (in the reflection principles themselves) other means of ascent”.<sup>285</sup> What Tennant has in mind here is that, gradually, new axioms and principles (owing to a process of repeated exposure?) appear to us to be self-evident. What Tennant has in mind is very much akin to the following Gödelian hypothesis of how one gains a grasp of indispensable mathematical axioms logically independent of those already held:

If one considers the development of a child, one notices that it proceeds in two directions: it consists on the one hand in experimenting with the objects of the external world and with its [own] sensory and motor organs, on the other hand in coming to a better and better understanding of language, and that means – as soon as the child is beyond the most primitive designating [of objects] – of the basic concepts on which it rests. With respect to the development in the second direction, one can justifiably say that the child passes through states of consciousness of various heights, e.g., one can say that a higher state of consciousness is attained when the child first learns the use of words, and similarly at the moment when for the first time it

---

<sup>283</sup> And, indeed, “one might...insist that *something* should explain why this is so”. (Ketland 2005, p. 80)

<sup>284</sup> And only in his short 2005 reply to Ketland 2005.

<sup>285</sup> Tennant 2005, p. 96.

understands a logical inference...a systematic and conscious advance in the second direction will...far exceed the expectations one may have a priori. In fact...even without the application of a systematic and conscious procedure, but entirely by itself, a considerable further development takes place in the second direction, one that transcends “common sense”. Namely, *it turns out that in the systematic establishment of the axioms of mathematics, new axioms, which do not follow by formal logic from those previously established, again and again become evident.* (Gödel 1961 (1995), pp. 383-385. Emphasis added.)

Tennant is submitting something precisely along these Gödelian lines: that reflection principles are examples of “new axioms that become evident” owing to one’s having achieved a “higher state of consciousness”, thereby negating any recourse to a substantial theory of truth. This illustrates, as Ketland points out, what Bertrand Russell once called “the advantages of theft over honest toil”. It seems then, according to Tennant, that the deflationist can assume principles without argument.

Ketland and Shapiro (and unbeknownst to Ketland at the time of writing his original 1999 paper, also Feferman in 1991) all deal with the issue of how strong the justificatory resources are required to be in proving global reflection principles such as ‘All theorems of  $PA$  are true’. The plan common to all was to use truth-theoretic extensions – this being all the more pertinent for deflationism given its insistence on the logical equipollence of theory  $PA$  with the claim ‘All theorems of  $PA$  are true’. In demonstrating that conservatively-extending truth theories are inadequate to prove global reflection principles, Ketland reasonably concludes that deflationary theories of truth tell much less than the whole truth about truth. As he points out, his argument deals with deductive content. Tennant’s 2002 paper deals with the non-deductive justificatory resources needed for apprehending the truth of the Gödel sentence  $G_{PA}$ , and so does not answer the point raised by Ketland at all. We must, therefore, resist the conclusion Tennant draws, namely that deflationist strategies are available for ‘reflecting’ upon essentially incomplete formal systems, and accept that, as with Hilbert’s programme, Gödelian considerations crushingly demonstrate the unworkability of the deflationary programme of deflating the concept of truth.

### 5.3 Appendix

To see how a proof of consistency for a formal system encoding abstract techniques establishes conservativity, let  $R$  denote a formal system encoding ‘real’ statements with finitistic proofs, and let  $I$  denote an ideal system involving abstract reasoning. Let  $\phi$  be the statement  $\forall x(fx = gx)$ , where the quantifier ranges over  $\mathbb{N}$  and  $f$  and  $g$  are primitive recursive functions. Let us assume a standard Gödel coding of the language of  $I$  in  $R$ . Now suppose  $I \vdash \phi$ . Hence, there is a number  $d$  such that (where  $Pr(u, v)$  is the primitive recursive predicate ‘ $u$  is the code number of a proof in  $I$  of the formula with code number  $v$ ’):

$$R \vdash Pr(d, \ulcorner \phi \urcorner), \text{ where } \ulcorner \phi \urcorner \text{ is the code for } \phi.$$

The following can be established (Smorynski 1977, p. 824):

$$R \vdash fx \neq gx \rightarrow Pr(c(x), \ulcorner \neg \phi \urcorner), \text{ for some derivation } c(x) \text{ depending on } x.^{286}$$

Now suppose  $R$  proves the consistency of  $I$ . It follows that:

$$R \vdash \neg[Pr(d, \ulcorner \phi \urcorner) \wedge Pr(c(x), \ulcorner \neg \phi \urcorner)]$$

$$\text{Hence, } R \vdash \neg Pr(c(x), \ulcorner \neg \phi \urcorner)$$

$$\text{Hence, } R \vdash fx = gx$$

$$\text{Hence, } R \vdash \forall x(fx = gx) \text{ by universal generalization.}$$

---

<sup>286</sup> This is because  $fx \neq gx$  for any given  $x$  is a finitary statement and hence provable in  $R$ , and since  $R$  is included in  $I$  it’s provable in  $I$  too. And since  $I$  contains logic,  $I$  then proves  $\exists x(fx \neq gx)$  and hence proves  $\neg \forall x(fx = gx)$ .

# CHAPTER 6

## CONCLUSION

That the elements of a picture are related to one another in a determinate way represents that the things are related to one another in the same way.

**Ludwig Wittgenstein, *Tractatus Logico-Philosophicus***

The question ‘What is truth?’ invites us to determine what the concept of truth consists in. It does not invite us to identify those qualities, other than truth, which belong to whatever is true.<sup>287</sup> To understand the question as inviting us to do the latter (as we learned in Chapter 2) is to err as the epistemicist errs. This is made starkly apparent in the following account of James’ instrumentalist view:

Our account of truth is an account of truths in the plural, of processes of leading, realized *in rebus*, and having only this quality in common, that they *pay*.  
(James 1907a, p. 104)

The trivial sort of correspondentism expressed in the *T*-schema, and one acting as a minimal constraint of adequacy, is not implied by any epistemic theory whatsoever. It is, however, implied by two rival theories: the correspondence theory and the deflationary theory. The fundamental difference between the two views lies in their differing attitude to the nature of the enquiry undertaken here. Inherent in the correspondence formula is, so the correspondentist maintains, an *analysis* of truth, one that identifies the nature of truth and endows it with nontrivial properties. In opposition, deflationists maintain that no such analysis is possible because to say that truth has a ‘nature’ is a misnomer, and one which has misled many into investigating what its features might be; there is on this view no philosophical value added to asking questions like ‘What is (the nature of) truth?’ for there is no nature to truth to be discovered. As mentioned above, these two views do accept that to say of a proposition that it is true is to say that the proposition corresponds with fact.<sup>288,289</sup> But the *T*-schema is just a (true) biconditional nowhere opting in favour of one or the

---

<sup>287</sup> Cf. the quote from Russell’s *Philosophical Essays* used as the epigram for Chapter 3.

<sup>288</sup> We saw in 4.2.1 and in 4.2.3 how the illocutionary theory of Strawson’s and the prosententialism of G.C.B. make no appeal to the *T*-schema whatsoever. To say ‘*p* is true’ is to say, according to Strawson, something like ‘Hooray to *p*’ and, according to prosententialism, the sentence ‘*p* is true’ stands in for the sentence denoted by *p* just as ‘she’ does for Mary in ‘Mary had a little lamb, but she preferred armadillos’. If no appeal is made to the *T*-schema in explicating truth then the minimal constraint found in the folk theory is not met. Both the performative and the prosentential theories are (rightly) not considered mainstream deflationary theories of truth.

<sup>289</sup> This is supported by the following deflationist’s remark: “*It is indeed undeniable that whenever a proposition or an utterance is true, it is true because something in the world is a certain way – something typically external to the proposition or utterance.*” (Horwich 1998a, p. 104).



other view; it is entirely neutral on that account. It features as a sort of explanandum (which is why it is a triviality acceptable to all) and according to Tarski's (and Kripke's) theory of a relation between a language and an interpretative structure, the biconditional appears only as a (low-level) deductive consequence. Tarski himself emphasized that the biconditional is not a definition of truth; rather it is an adequacy condition for such a definition; *T*-biconditionals are to be derived from, and explained by, a definition linking expressions to their interpretation in the world. If the object language is included in the metalanguage, we can get Tarski's standard example:

‘snow is white’ is true if and only if snow is white

which expresses a relation (of the particular strong correspondence type) between a sentence of the (tacitly) indexed object language and a fact. Certainly there are true instances such as:

‘Phlogiston is given off during combustion’ is true if and only if Phlogiston is  
given off during combustion

which seem to commit us to a particularly peculiar fact, *viz.* the existence of Phlogiston,<sup>290</sup> but equivalence statements can be true even when both sides are false. In cases where the object language is not included in the metalanguage – where, for example, *L* is German and *ML* is English – then instances such as:

‘Schnee ist weiss’ is true if and only if snow is white

are to be understood in such a way that the terms ‘Schnee’ and ‘weiss’ are interpreted in the world to denote those entities that are denoted by ‘snow’ and ‘white’ in English.

According to deflationism (the ‘empty theory’), on the other hand, the biconditional itself is all one knows, and – according to that theory – all one in fact needs to know. But as we saw, this view has unacceptable consequences. For example, it seems an obvious adequacy condition on an acceptable theory of truth that it should be able to prove that if a set of sentences is true then the set is consistent. But the *T*-schema combined with Gödel's second incompleteness theorem tells us that the consistency of the Peano axioms is not a consequence of them (if they are in fact consistent) while the truth of each is. If anything should be taken as a reductio of deflationism, this surely should.

---

<sup>290</sup> This criticism is made in Jennings 1987.

We now summarize the findings of this work:

We conclude that the traditional problems associated with the correspondence theory are ungrounded. That the fact-operator ‘*it is a fact that...*’ says the same thing as the truth-operator ‘*it is true that...*’ is not taken to imply that a fact is nothing more than a reification of a style of speaking (or, at any rate, we deny that this observation implies more is not involved). A rudimentary folk, fact-based semantics would naturally be misleading. This is the insight we gain from our examination of the slingshot argument. That the argument is valid should not disturb our state of *ataraxia* but force us to recognize the need to evolve from coarse-grained theories to finer-grained ones that are scientifically more sophisticated. Indeed, this is also the lesson to be learnt from the Liar paradox. Tarski’s development of *truth-in-a-model* placed the correspondence theory in a clearly delimited formal setting. This accomplished two things. First, the threat posed by the Liar is removed since the theory’s consistency is guaranteed by Tarski’s theorem. This theorem tells us that the paradox is not reproducible since ‘is true’ is not definable in the object language; there is, therefore, nothing in the nature of truth itself that generates paradox – it arises, if you like, because we attempted to do something akin to dividing by 0, i.e. thinking that, since there is a word ‘true’ in English, English is a semantically closed language in which the central properties of truth are maintained. Tarski showed that you cannot have this together with bivalence (Kripke in his celebrated work on truth showed you could have semantical closure, but at the cost of bivalence).<sup>291</sup> Second, Tarski’s work allows us to see exactly how true sentences correspond to facts: true sentences are homomorphic images of facts, i.e. a true sentence represents, in a form-preserving manner, the truth-making facts in it. To see this, examine the clauses of a Tarskian truth definition: (a) basis clause:–  $R(a_1, \dots, a_n)$  is true in  $I$  iff  $I(R)(I(a_1), \dots, I(a_n))$ , where  $I(x)$  is the denotation in the structure  $I$  of the linguistic item  $x$ ; (b) induction clauses:– (b<sub>i</sub>)  $\neg A$  is true in  $I$  iff  $\neg(A$  is true in  $I)$ ; (b<sub>ii</sub>)  $A \vee B$  is true in  $I$  iff  $A$  is true in  $I$  or  $B$  is true in  $I$ ; (b<sub>iii</sub>)  $\forall x A(x)$  is true in  $I$  iff for all  $I(a)$ ,  $A[I(a)]$  is true in  $I$ . On the right hand side of each clausal biconditional you have a condition with exactly the same logical form as the sentence on the left (and, with no mention of truth, this forestalls Frege’s objection), employing the same atomic formula structure for atomic sentences and the same

---

<sup>291</sup> Indeed, not only can ‘ordinary’ negation not be defined (for then a Liar sentence could be reconstructed) in Kripke’s semantically closed language  $L$  containing its own truth predicate, but the sentence ‘ $\lambda$  is not true’, where  $\lambda$  is a Liar-sentence of  $L$ , itself lacks a truth-value even though it is intuitively true. Such statements can only be evaluated correctly (according to intuitive criteria which nonetheless seem very compelling) in the classical metalanguage. But then, as Kripke himself admitted, we are still in need of a Tarskian metalanguage to make intuitive sense of what goes on in the supposedly semantically closed language  $L$ .

connectives and quantifiers for the compound sentences; the corresponding facts are built up recursively matching the functional composition portrayed in the sentence-structure.<sup>292</sup>

The popular objections raised against this theory are not well founded. We saw, in Chapter 3, how the ‘list-like’ objection popularized by Field, which claims that all that is achieved are definitions of the form ‘ $\lambda$  is true <sub>$L$</sub> ’ (and not crucially of the form ‘ $\lambda$  is true-in- $L$ ’ for variable  $L$ ), misses the point that the model-theoretic definition provides us with a uniform definition applying to *all* structures capable of interpreting  $L$ ’s descriptive vocabulary. Indeed, contrary to the objection that Tarski’s work dealt exclusively with uninterpreted, formalized languages with nothing of interest to say about natural languages, the formalized languages Tarski employed are best viewed as simple models sharing sufficiently many features with natural languages so as to be capable of serving as experiments with which to test out certain claims and properties. Tarski understood this well. As he himself said: “the only formalized languages that seem to be of real interest are those which are fragments of natural languages.”<sup>293</sup>

Tarski’s theory has, of course, been famously criticized by Kripke (and others; since Kripke is the author of the most serious objections I shall restrict my remarks to him), precisely on the ground that it does not model natural language well: indeed very badly in some important ways. In particular, qua model of English truth-discourse, the Tarski theory implicitly stratifies truth-in-English into a potentially infinite hierarchy<sup>294</sup> of distinct truth predicates such that truth at level  $\alpha$  has in its scope only the truth predicates on levels beneath  $\alpha$ , on pain of ushering in the Liar paradox. But English appears to have a single truth predicate, not a numberless hierarchy, within whose scope moreover it is quite consistent to have sentences containing that predicate. Kripke 1975 has supplied a famous example with his Nixon-Dean imaginary dialogue (Nixon was President of the USA at the time, with Watergate a recent scandal). Dean states ‘All Nixon’s

---

<sup>292</sup> This is how Tarski’s theory meets Wittgenstein’s (early) view that true sentences correspond to facts by being pictures of them. Both understood that the logical structure of sentences is in every case a functional composition of corresponding names of Boolean functions. Thus,  $\wedge$  names the Boolean meet,  $\vee$  names the Boolean join,  $\neg$  names complementation  $\sim$ , and  $\forall$  and  $\exists$  name the infinitary Boolean meet and join over all instances. The truth-table rules just ensure that the truth-value in the Boolean algebra  $\{0,1\}$  is given by the composite Boolean function mirrored in the structure of the proposition. For example, consider the formula ‘ $p \wedge \neg(q \vee r)$ ’. Using 1 to symbolize ‘true’ and 0 to symbolize ‘false’, suppose  $v(p) = 1$ ,  $v(q) = 0$  and  $v(r) = 0$  under a valuation  $v$  of the generators  $p$ ,  $q$  and  $r$ . ‘ $p \wedge \neg(q \vee r)$ ’ is true under the valuation  $v$ , i.e.  $v(p \wedge \neg(q \vee r)) = 1$ , just in case the Boolean function  $1 \wedge \neg(0 \vee 0)$  has the value 1 in the Boolean algebra  $\{0,1\}$ , which it has. More generally,  $v(p \wedge \neg(q \vee r))$  is the value of the function  $v(p) \wedge \neg(v(q) \vee v(r))$ . This can be represented by the kind of commutative diagram given in 3.7 (iv). The rigorous and general development of this algebraic approach is provided in Rasiowa & Sikorski 1968.

<sup>293</sup> Tarski 1969, p. 68.

<sup>294</sup> In the model, actually infinite, up to the first nonrecursive ordinal (see Halbach 1999b).

assertions about Watergate are false' (1), but among Nixon's assertions about Watergate is the following: (2) 'Everything Dean says about Watergate is true'. But suppose that Dean had made at least one unequivocally true assertion about Watergate (e.g. 'Watergate is Watergate'). In that case Nixon's statement (2) is clearly false, and if all Nixon's other statements about Watergate are false then Dean's assertion (1) is true. There is apparently nothing paradoxical or inconsistent in the way we arrive at these evaluations: indeed, given the supposed facts we are pretty much compelled to them. Yet in being compelled to them we use the fact that (1) and (2) lie in each other's scope, and so the evaluations are precluded within the Tarski model (which Kripke calls the "orthodox approach"). As Kripke remarks, it is possible to construct all sorts of similarly intuitive and consistent truth-evaluations which are difficult if not impossible to accommodate with the Tarski approach (Kripke 1975 (1984), p. 60).

This seems to be a profound objection to Tarski's theory, at any rate insofar as the latter claims to model ordinary, reasonable truth-discourse; if it does not then it is just a formal model of purely logico-mathematical, but not philosophical, interest. Kripke's celebrated response was to construct an alternative formal model to Tarski's in which a single predicate *Tr* is included in the extralogical vocabulary of the formal (in fact, first order) language *L*,<sup>295</sup> whose intended interpretation is a correspondentist truth à la Tarski but with the difference that the semantics allow *Tr* to be undefined for certain of its arguments, including all Liar-type sentences constructible in *L*: Tarski's theorem, proved by a Liar-type reductio, tells us that *Tr* can't be a total predicate. The existence of truth-value gaps of course necessitates a revision of standard logic, since it is based on bivalence, but Kleene had several years before, and for a quite different purpose (to construct a logic of computable predicates and functions), exhibited a three-value logic, in fact two, the so-called weak and strong systems, the latter of which Kripke used for his own model of truth.

Kripke's theory has itself come in from serious criticism, not least by himself, not only for being unable to accommodate intuitive judgements about truth that Tarski's theory can accommodate (for example, as we have already noted, that the Liar sentence is not true), but also failing to eradicate stratification: its construction is carried out in a classical metalanguage. But these matters, and the formal details of Kripke's theory, don't concern us here: from the point of view of the present

---

<sup>295</sup> *L* can either be a language in which self-reference is defined via a Gödel-type encoding and the construction of a diagonal function, or simply be such that its standard interpretation contains all its sentences as a subdomain, in the manner of Martin and Woodruff 1975. Martin and Woodruff's language also contains its own truth predicate, like Kripke's, and the authors use Zorn's Lemma to prove a fixed-point theorem corresponding to Kripke's. Coincidentally, Martin and Woodruff's seminal paper appeared in the same year as Kripke's, though Kripke's is better-known because of its accompanying philosophical discussion which included a remarkable statement of some of the serious deficiencies of his own theory.

discussion, the important point to note is that Kripke's theory is just as correspondentist as Tarski's: only the underlying logic is different. As to how well a formal model can model all the intuitive judgements we might want to make, about truth-discourse, or the set of all sets, the lesson of the semantical paradoxes no less than the set-theoretical paradoxes is that some of these 'intuitions' have to be given up on pain of inconsistency. Kripke's theory may 'correct' Tarski's in some ways relative to these intuitions, but at the inevitable cost of failing to be 'correct' relative to others. No model will be 'correct' relative to all. Whether being able to accommodate the Nixon-Dean exchange is more important than having a sense of negation that allows us to say that the Liar sentence is not true is a matter of opinion. Mine is that it is not.

The most popular objection to Tarski's theory comes from deflationism. While the Tarskian views the *T*-schema as a consequence of his theory which defines truth in terms of attributing non-linguistic 'worldly' properties to non-linguistic 'worldly' entities,<sup>296</sup> most deflationists view the *T*-schema as conceptually basic; that is, no logical excavations taking us deeper are needed (or, indeed, are possible) to understand what truth consists in. On this opposing view the *T*-schema underwrites truth's ability to state useful generalizations and make blind ascriptions/endorsements but it is entirely to this logico-linguistic facility that truth owes its existence; that truth is usefully an artefact of bookkeeping habits<sup>297</sup> does *not* imply, for the deflationist, that truth exists beyond this function. This *means* that deflationism about truth can be represented by the idea that 'adding' truth to a (non-semantical) theory of the world – let us use *PA* – adds no new informational content, or garners no new (mathematical) facts.<sup>298</sup> It is in this sense that deflationary truth is insubstantial.<sup>299</sup> Technically, this amounts to the claim that this theory *conservatively extends PA*. However, one's acceptance of the base theory (here, *PA*) commits one to an acceptance of further statements in the language of the base theory (and one of these for *PA* is its Gödel sentence  $G_{PA}$ ).<sup>300</sup> And this commitment, as reported in Chapter 5, uses the notion of truth, thus allying truth with its non-conservativeness i.e. substantiality. The ability to recognize the truth of the Gödel sentence  $G_{PA}$  involves a theory of truth, i.e. Tarski's, significantly transcending the deflationary theory.<sup>301</sup> This

---

<sup>296</sup> Cf. Sher 1999, pp. 134-135.

<sup>297</sup> I borrow this phrase from Simon Blackburn.

<sup>298</sup> Cf. Shapiro 1998, p. 497 and Ketland 1999, p. 79.

<sup>299</sup> If truth is not redundant in this sense, then it is difficult to see what, alternatively, could be understood by the claim that truth is insubstantial.

<sup>300</sup> Ketland calls this a "*conditional epistemic obligation*" (Ketland 2005, p. 80).

<sup>301</sup> Cf. Ketland 1999, p.88.

means that we are, after all, entitled to place instead of Ramsey's ladder Tarski's sturdier replacement, professing greater insights as we ascend.

The propositions – such as ‘All theorems of *PA* are true’<sup>302</sup> – which form part of the deflationary package are *not provable* in the conservatively extending theory. This means that the proposition ‘All theorems of *PA* are true’ and an acceptance of *PA* are not equivalent, i.e. there are commitments that logically follow from the first proposition that *do not* follow from a mere acceptance of *PA*. The reflective proposition is *deductively stronger* contrary to what the deflationist contends; there is a *logical difference* in accepting *PA* and accepting *PA*'s soundness statement.<sup>303</sup> Hence, the deflationary claim that the instances of the *T*-schema entail all there is to be said about truth, that all the facts about truth can be explained solely on the basis of them, is false. These are only attainable with a robust, substantial, non-conservative theory of truth,<sup>304</sup> which is the theory devised by Tarski and defended here.

---

<sup>302</sup> Indeed, it is a condition of adequacy on any theory of truth that it be able to prove the equivalence of a (possibly infinitely axiomatized) theory and its truth (*cf.* Ketland 1999, p. 90).

<sup>303</sup> Hence, to argue as Tennant 2005 does that the Gödel sentence  $G_{PA}$  can follow from the soundness of *PA* in a deflationarily licit way is not relevant to the dialectic here. What is important, however, is the manner in which we arrive at the soundness claim. The justificatory resources needed to prove truth-theoretic generalizations requires, contrary to deflationary claims, more than the instances of the *T*-schema.

<sup>304</sup> A result that Tarski had already established in his 1936 *locus classicus* (Tarski 1936a (1983), p. 274).

# BIBLIOGRAPHY

**ACTON, H.B.** 1935, "The Correspondence Theory of Truth", *Proceedings of the Aristotelian Society*, **35**, pp. 177-194.

**ALSTON, W.P.** 1996, *A Realist Conception of Truth*, Cornell University Press, Ithaca, New York.

**AQUINAS, T.** 1952-1954, *Questiones Disputatae De Veritate*, translated by R.W. Mulligan, J.V. McGlynn & R.W. Schmidt as *Truth*, Regnery, Chicago.

## ARISTOTLE

- *Metaphysics*. In *The Complete Works of Aristotle: The Revised Oxford Translation*, edited J. Barnes. Princeton University Press, New Jersey, 1984.
- *Categories*. In *The Complete Works of Aristotle: The Revised Oxford Translation*, edited J. Barnes. Princeton University Press, New Jersey, 1984.

## ARMOUR-GARB, B.

- 2005a (editor with J.C. Beall), *Deflationary Truth*, Open Court Readings in Philosophy.
- 2005b (editor with J.C. Beall), *Deflationism and Paradox*, Oxford University Press, Oxford.

## AUSTIN, J.L.

- 1950, "Truth", *Proceedings of the Aristotelian Society*, supplementary volume, **24**, pp. 111-129. Reprinted in Austin, J.L. 1961 (1970), pp. 117-133. Also reprinted in Pitcher, G. (ed.) 1964, pp. 18-31 and in Blackburn, S. & Simmons, K. (eds.) 1999, pp. 149-161.
- 1954, "Unfair to Facts", in Austin, J.L. 1961 (1970), pp. 102-122. Reprinted in Blackburn, S. & Simmons, K. (eds.) 1999, pp. 183-199.
- 1961, *Philosophical Papers*, Oxford University Press, Oxford. Second edition 1970.

## AYER, A.J.

- 1936, *Language, Truth and Logic*, Pelican Books, London. Latest edition 1971.
- 1953, "Truth", *Revue Internationale de Philosophie*, **7**, pp. 183-200.
- 1963, "Truth" in his *The Concept of a Person and Other Essays*, pp. 162-187, St. Martin's Press, New York, First edition 1963. Reprinted edition 1964, Macmillan, London.

**BARWISE, J.**

- 1981a (with J. Perry), "Semantic Innocence and Uncompromising Situations" in French, P.A., Uehling (Jr.), T.E. & Wettsteing, H.K. (eds.) 1981, *Midwest Studies in Philosophy: The Foundations of Analytic Philosophy*, Volume VI, University of Minnesota Press, Minneapolis, pp. 387-403.
- 1981b (with J. Perry), "Situations and Atitudes", *Journal of Philosophy*, **78**, pp. 668-691.
- 1987 (with J. Etchemendy), *The Liar: An Essay on Truth and Circularity*, Oxford University Press, Oxford.

**BELNAP, N.** 1973, "Restricted Quantification and Conditional Assertion" in Leblanc, H. (ed.), *Truth, Syntax and Modality*, pp. 48-75, North Holland, Amsterdam.

**BENACERRAF, P. & PUTNAM, H.** (eds.) 1964, *Philosophy of Mathematics: Selected Readings*. Revised second edition. Cambridge University Press, Cambridge.

**BERKELEY, G.**

- 1710, *Principles Human Knowledge*, Oxford World Classics 1999.
- 1734, *Three Dialogues*, Oxford World Classics 1999.

**BLACKBURN, S.W.**

- 1984, *Spreading the Word: Groundings in the Philosophy of Language*, Oxford University Press, Oxford.
- 1998, *Ruling Passions: A Theory of Practical Reasoning*, Oxford University Press, Oxford.
- 1999 (edited with K. Simmons) *Truth*, Oxford University Press, Oxford.
- 2005, *Truth: A Guide for the Perplexed*, Penguin, London.

**BLANSHARD, B.** 1941, *The Nature of Thought*, George Allen & Unwin, London.

**BOVENS, L. & HARTMANN, S.** 2003, *Bayesian Epistemology*, Oxford University Press, Oxford.

**BRADLEY, F.H.**

- 1907, "On Truth and Copying" in Bradley, F.H. 1914, pp. 107-126.
- 1909, "On Truth and Coherence" in his *Writings on Logic and Metaphysics*, edited by J.W. Allard & G. Stock, Clarendon Press, Oxford, F.H. 1994, pp. 202-218.



- 1914, *Essays on Truth and Reality*, Oxford University Press, Oxford.

**BRANDOM, R.B.** 1994, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Harvard University Press, Cambridge, Massachusetts.

**BRENTANO, F.C.**

- 1956, *Die Lehre vom richtigen Urteil*, Francke Verlag, Bern.
- 1970, *Versuch über die Erkenntnis*, Meiner, Hamburg.

**CANDLISH, S. & DAMNJANOVIC, N.** 2007, "A Brief History of Truth" in D. Jacquette (ed.) 2007, *Philosophy of Logic*, North-Holland, Amsterdam, pp. 227-324.

**CARNAP, R.**

- 1939, "Foundations of Logic and Mathematics" in *International Encyclopedia of Unified Science*, **1**, University of Chicago Press, Chicago.
- 1942, *Introduction to Semantics*, Harvard University Press, Cambridge, Massachusetts.
- 1947, *Meaning and Necessity: A Study in Semantics and Modal Logic*, University of Chicago Press, Chicago. Second edition 1956.
- 1962, *The Logical Foundations of Probability*, University of Chicago Press, Chicago.
- 1963, "Replies and Systematic Expositions" in P.A. Schilpp (ed.) 1963, *The Philosophy of Rudolf Carnap*, Open Court, La Salle, pp. 859-1013.

**CHANG, C. & KEISLER, H.** 1973, *Model Theory*, North-Holland, Amsterdam.

**CHISHOLM, R.M.** 1977, *Theory of Knowledge*, Prentice-Hall, Englewood Cliffs, New Jersey, Second Edition.

**CHURCH, A.**

- 1943, "Review of Carnap's *Introduction to Semantics*", *Philosophical Review*, **52**, pp. 298-304.
- 1956, *Introduction to Mathematical Logic: Volume 1*, Princeton University Press, Princeton.

**DAVID, M.**

- 1994, *Correspondence and Disquotation*, Oxford University Press, Oxford.

- 2005, "The Correspondence Theory of Truth", entry in *The Stanford Encyclopaedia of Philosophy* (edited by E.N. Zalta).

**DAVIDSON, D.**

- 1967a, "Truth and Meaning", *Synthese*, **17**, pp. 304-323. Reprinted in Davidson, D. 1984, pp. 17-36.
- 1967b, "Causal Relations", *Journal of Philosophy*, **64**, pp. 691-703. Reprinted in Davidson, D. 2001, *Essays on Actions and Events*, Clarendon Press, Oxford, Second Edition.
- 1969, "True to the Facts", *Journal of Philosophy*, **66**, pp. 748-764. Reprinted in Davidson, D. 1984, pp. 37-54.
- 1977, "The Method of Truth in Metaphysics" in P.A. French, T.A. Uehling (Jr.) & H.K. Wettstein (eds.), 1977, *Midwest Studies in Philosophy 2: Studies in the Philosophy of Language*, University of Minnesota Press, Minnesota. Reprinted in Davidson, D. 1984a, pp. 199-214.
- 1984, *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford.
- 1990, "The Structure and Content of Truth", *Journal of Philosophy*, **87**, pp. 279-328.
- 1996, "The Folly of Trying to Define Truth", *Journal of Philosophy*, **93**, pp. 263-278. Reprinted in Blackburn & Simmons (eds.) 1999, pp. 308-322. Also reprinted in Lynch, M.P. (ed.) 2001, pp. 623-640.
- 1999, "The Centrality of Truth" in Peregrin, J. (ed.) 1999a, pp. 105-115.

**DESCARTES, R.** 1639, "Letter to Mersenne: 16 October 1639", in *The Philosophical Writings of Descartes*, Volume **3**, 1991, pp. 138-140, Cambridge University Press, Cambridge

**DEWEY, J.** 1911, "The Problem of Truth", in *John Dewey, The Middle Works 1899-1924: Journal Articles, Book Reviews, Miscellany in the 1910-1911 period & HOW WE THINK*, Volume **VI**, pp. 12-68, Southern Illinois University Press, Carbondale.

**DOUVEN, I. & HINDRINKS, F.** 2005, "Deflating the Correspondence Intuition", *Dialectica*, **59**, pp. 315-329.

**DRETSKE, F.** 1969, *Seeing and Knowing*, Routledge, London

**DUCASSE, C.J.** 1940, "Propositions, Opinions, Sentences, and Facts", *Journal of Philosophy*, **37**, pp. 701-711.

**DUMMETT, M.A.**

- 1959, "Truth", *Proceedings of the Aristotelian Society*, **59**, pp. 141-162. Reprinted in Picher, G. (ed.), pp. 93-111, in Dummett, M.A. 1978, pp. 1-21, and in Strawson, P.F. (ed.) 1973, pp. 49-68.
- 1978, *Truth and Other Enigmas*, Harvard University Press, Cambridge, Massachusetts.

**DUNN, J.M.**

- (with A. Gupta) (eds.) 1990, *Truth or Consequences: Essays in Honour of Nuel Belnap*, Kluwer, Dordrecht.
- (with G. Hardegree) 2001, *Algebraic Methods in Philosophical Logic*, (Oxford Logic Guides, Volume 41), Clarendon Press, Oxford.

**EVANS, G. & McDOWELL, J.** (eds.) 1976, *Truth and Meaning: Essays in Semantics*, Clarendon Press, Oxford.

**EWING, A.C.** 1934, *Idealism: A Critical Survey*, George Allen & Unwin, London.

**FEFERMAN, S.** 1991, "Reflecting on Incompleteness", *Journal of Symbolic Logic*, **56**, pp. 1-49.

**FIELD, H.H.**

- 1972, "Tarski's Theory of Truth", *Journal of Philosophy*, **69**, pp. 347-375. Reprinted in Field, H.H. 2001a, pp. 3-29. Also reprinted in Lynch, M.P. (ed.) 2001, pp. 365-396 and in Schmitt, F.F. 2004a, pp. 152-179.
- 1980, *Science Without Numbers: A Defence of Nominalism*, Princeton University Press, Princeton, New Jersey.
- 1986, "The Deflationary Conception of Truth" in Macdonald G. & Wright C. (eds.) 1986, pp. 55-117.
- 1989, *Realism, Mathematics and Modality*, Blackwell, Oxford.
- 1994a, "Deflationist Views of Meaning and Content", *Mind*, **103**, pp. 249-284. Reprinted in Blackburn & Simmons (eds.) 1999, pp. 351-391, in Schmitt, F.F. 2004a, pp. 253-294 and (with a postscript) in Field, H.H. 2001a, pp. 104-156.
- 1994b, "Disquotational Truth and Factually Defective Discourse", *Philosophical Review*, **103**, pp. 405-452. Reprinted in Field, H.H. 2001a, pp. 222-258.
- 1998, "Deflating the Conservativeness Argument", *Journal of Philosophy*, **95**, pp. 533-540.

- 2001, *Truth and the Absence of Fact*, Clarendon Press, Oxford.

**FRENCH, P.A., UEHLING (Jr.), T.E. & WETTSTEIN, H.K.** 1988, *Midwest Studies in Philosophy: Realism and Antirealism*, Volume **XII**, University of Minnesota Press, Minneapolis.

**FREGE, G.**

- 1892, “Über Sinn und Bedeutung”, *Zeitschrift für Philosophie und Philosophische Kritik*, **100**. Translated as “On Sense and Reference” in A.W. Moore (ed.) 1993, *Meaning and Reference*, pp. 23-42, Oxford University Press, Oxford. Also translated as “On Sense and Meaning” in Geach, P. & Black, M. (eds.) 1980, pp. 56-78. Both translations provided by Max Black.
- 1897, *Logic*, in Hermes, H., Kambartel, F. & Kaulbach, F. (eds.) 1979.
- 1906 (or earlier), *17 Key Sentences on Logic*, in Hermes, H., Kambartel, F. & Kaulbach, F. (eds.) 1979.
- 1911, “Der Gedanke”, in Frege, G. 1979. Translated as “The Thought: A Logical Inquiry” by A.M. Quinton and M. Quinton and appeared in 1956, *Mind*, **65**, pp. 289-311. Reprinted in Blackburn & Simmons (eds.) 1999, pp. 85-105, and in Strawson, P.F. (ed.) 1973, pp. 17-38.
- 1979, *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel/Posthumous Writings*, edited by Hans Hermes, Friedrich Kambartel, Friedrich Kaulbach, with the assistance of Gottfried Gabriel and Walburga Rödding. Translated by Peter Long, Roger White, with the assistance of Roger Hargreaves. Volume 1, Selections, University of Chicago Press, Chicago.

**FROST-ARNOLD, G.** 2004, “Was Tarski’s Theory of Truth Motivated by Physicalism?”, *History and Philosophy of Logic*, **25**, pp. 265-280.

**GARCÍA-CARPINTERO, M.** 1998 (with M. Pérez Otero), “Davidson, Correspondence Truth and the Frege-Gödel-Church Argument”, *History and Philosophy of Logic*, **19**, pp. 63-81.

**GEACH, P.**

- 1960, “Ascriptivism”, *Philosophical Review*, **69**, pp. 221-225.
- 1967, “Intentional Identity”, *Journal of Philosophy*, **64**, pp. 627-632.
- 1980 (editor with M. Black), *Translations from The Philosophical Writings of Gottlob Frege*, Blackwell, Oxford. First edition 1952.

**GÖDEL, K.**

- 1931, “Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I”, *Monatshefte für Mathematik und Physik*, **38**, pp. 173-198. The English translation “On Formally Undecidable Propositions in *Principia Mathematica* and Other Systems I” first appeared in 1951 and is reprinted in van Heijenoort, J. (ed.) 1967, pp. 596-616.
- 1944, “Russell’s Mathematical Logic” in Schilpp, P.A. (ed.) 1951, *The Philosophy of Bertrand Russell*, Tudor Publishing, New York. Reprinted in Benacerraf & Putnam (eds.) 1983, pp. 447-469.

**GRAYLING, A.** 1997, *An Introduction to Philosophical Logic*, Blackwell, Oxford.

**GRIM, P.** 1991, *The Incomplete Universe*, MIT Press, Cambridge, Massachusetts.

**GROVER, D.L.**

- 1975 (with J.L. Camp (Jr.) & N.D. Belnap (Jr.)), “A Prosentential Theory of Truth”, *Philosophical Studies*, **27**, pp. 73-125. Reprinted in Grover, D. 1992, pp. 70-120 and (sections 1 to 3 only) in Schmitt, F.F. 2004a, pp. 219-252.
- 1990, “On Two Deflationary Truth Theories” in Dunn, J.N. & Gupta, A. (eds.) 1990, pp. 1-17.
- 1992, *A Prosentential Theory of Truth*, Princeton University Press, Princeton.
- 2001, “The Prosentential Theory: Further Reflections on Locating Our Interest in Truth”, in Lynch, M.P. (ed.) 2001, pp. 505-526.
- 2002, “On Locating Our Interest in Truth” in Schantz, R. (ed.) 2002, pp. 120-132.

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

**HALBACH, V.**

- 1999a, “Disquotationalism and Infinite Conjunctions”, *Mind*, **108**, pp. 1-22.
- 1999b, “Conservative Theories of Classical Truth”, *Studia Logica*, **62**, pp. 353-370.
- 2000a, “Disquotationalism Fortified” in A. Gupta & A. Chappuis (eds.) 2000, *Circularity, Definitions and Truth*, pp. 155-176, Indian Council of Philosophical Research, New Delhi.
- 2000b, “Truth and Reduction”, *Erkenntnis*, **53**, pp. 97-126.
- 2001a, “How Innocent is Deflationism?”, *Synthese*, **126**, pp. 167-194.

- 2001b, “Disquotational Truth and Analyticity”, *Journal of Symbolic Logic*, **66**, pp. 1959-1973.
- 2002, “Modalized Disquotationalism”, in V. Halbach & L. Horsten (eds.) 2002, *Principles of Truth*, pp. 75-101, Hänsel-Hohenhausen, Frankfurt.
- 2003, “Can We Grasp Consistency?” in E.J. Olsson (ed.), 2003, *The Epistemology of Keith Lehrer*, pp. 75-87, Kluwer, Dordrecht.
- 2004, *Semantics and Deflationism*, unpublished habilitation thesis.

**HALE, B. & WRIGHT, C.** (eds.) 1999, *A Companion to the Philosophy of Language*. Second edition (First edition, 1997), Blackwell, Oxford.

**HARE, R.M & RUSSELL, D.A.** (eds.) 1970a, *The Apology and Other Dialogues*, Volume I of *The Dialogues of Plato* series, B. Jowett (trans.), Hazell Watson & Viney, London.

**HARMAN, G.** 1971, “Substitutional Quantification and Quotation”, *Noûs*, **5**, pp. 213-214.

**HAZEN, A.P.** 1998, “Non-constructive Rules of Inference” in *The Routledge Encyclopaedia of Philosophy*, edited by E. Craig, Routledge, London.

**van HEIJENOORT, J.** (ed.) 1967, *From Frege to Gödel*, Harvard University Press, Cambridge, Massachusetts.

**HEMPEL, C.G.** 1935, “On the Logical Positivist’s Theory of Truth”, *Analysis*, **2**, pp. 49-59.

**HERMES, H., KAMBARTEL, F. & KAULBACH, F.** (eds.) 1979, *Gottlob Frege – Posthumous Writings*, University of Chicago Press, Chicago.

**HILL, C.S.** 2001, “The Marriage of Heaven and Hell: Reconciling Deflationary Semantics with Correspondence Intuitions”, *Philosophical Studies*, **104**, pp. 291-321.

**HINTIKKA, J.**

- 1975, “Carnap’s Heritage in Logical Semantics” in J. Hintikka (ed.), *Rudolf Carnap, Logical Empiricist*, D. Reidel, Dordrecht, pp. 217-242.
- 1996, *The Principles of Mathematics Revisited*, Cambridge University Press, Cambridge.

- 1997, *Lingua Universalis vs. Calculus Ratiocinator: An Ultimate Presupposition of Twentieth-Century Philosophy*, Kluwer, Dordrecht.

**HODGES, W.** 1986, "Truth in a Structure", *Proceedings of the Aristotelian Society*, **86**, pp. 135-152.

**HORWICH, P.**

- 1990, *Truth*, Blackwell, Oxford.
- 1998a, *Truth*. Second Edition. Clarendon Press, Oxford.
- 1998b, "The Minimalist Conception of Truth" in Blackburn, S. & Simmons, K. (eds.) 1999, pp. 239-264; equivalent to a revised edition of the postscript to Horwich 1998a.

**JAMES, W.**

- 1907a, *Pragmatism: A New Name for Some Old Ways of Thinking, The Works of William James*, Volume **I**, Harvard University Press, Cambridge, Massachusetts. Reprinted edition 1975.
- 1907b, "Pragmatism's Conception of Truth" in James, W. 1907a, pp. 197-236. Reprinted in Blackburn & Simmons (eds.) 1999, pp. 53-68.
- 1909, *The Meaning of Truth: A Sequel to Pragmatism' The Works of William James*, Volume **II**, Harvard University Press, Cambridge, Massachusetts. Revised edition 1975.

**JANSSEN, T.M.V.** 1997, "Compositionality" in J. Van Benthem & A. Ter Meulen (eds.) 1997, *Handbook of Logic and Language*, MIT Press, Cambridge, Massachusetts, pp. 417-473.

**JASTROW, J.** 1899, "The Mind's Eye", *Popular Science Monthly*, **54**, pp. 299-312.

**JENNINGS, R.C.** 1987, "Tarski: A Dilemma", *Inquiry*, **30**, pp. 155-172.

**JOACHIM, H.H.** 1906, *The Nature of Truth*, Oxford University Press, Oxford. Pages 72-79 are reprinted in Blackburn & Simmons (eds.) 1999, pp. 46-52.

**KALISH, D, MONTAGUE, R. & MAR, G.,** 1980, *Logic: Techniques of Formal Reasoning*, Harcourt Brace Jovanovich, New York.

**KANT, I.** 1781, *Critique of Pure Reason*, revised and expanded translation by J.M.D. Meiklejohn, edited by V. Politis 1993, Everyman, London.

**KATZ, J. & FODOR, J.** 1963, "The Structure of a Semantic Theory", *Language*, **39**, pp. 170-210.

**KAYE, R.** 1991, *Models of Peano Arithmetic* (Oxford Logic Guides, Volume 15), Oxford University Press, Oxford..

**KETLAND, J.J.**

- 1999, "Deflationism and Tarski's Paradise", *Mind*, **108**, pp. 69-94.
- 2000, "Conservativeness and Translation Dependent T-schemes", *Analysis*, **60**, pp. 319-328.
- 2005, "Deflationism and the Gödel Phenomena – Reply to Tennant", *Mind*, **114**, pp. 75-88.

**KIRKHAM, R.L.** 1992, *Theories of Truth: A Critical Introduction*, Bradford Books, MIT Press, Cambridge, Massachusetts.

**KITCHER, P.** 2002, "On the Explanatory Rôle of Correspondence Truth", *Philosophy and Phenomenological Research*, **64**, pp. 346-364. Reprinted in Schmitt, F.F. 2004a, pp. 197-215.

**KOTARBIŃSKI, T.**, 1929, *Gnosiology: The Scientific Approach to the Theory of Knowledge*, English translation from the Second Polish edition, 1966, Pergamon Press, Oxford.

**KOTLARSKI, H., KRAJEWSKI, S. & LACHLAN, A.** 1981, "Construction of Satisfaction Classes for Nonstandard Models", *Canadian Mathematical Bulletin*, **24**, pp. 283-293.

**KRIPKE, S.A.**

- 1975, "Outline of a Theory of Truth", *Journal of Philosophy*, **72**, pp. 690-716. Reprinted in Martin R.M. (ed.) 1984, pp. 53-81.
- 1976, "Is There a Problem with Substitutional Quantification?" in Evans, G. & McDowell, J. (eds.) 1976, pp. 325-419.

**KRÜGER, L.** 1995, "Has the Correspondence Theory of Truth Been Refuted? From Gottlob Frege to Donald Davidson", *European Journal of Philosophy*, **3**, pp. 157-173.



- KÜNNE, W.** 2003, *Conceptions of Truth*, Clarendon Press, Oxford.
- LAUDAN, L.** 1990, *Science and Relativism: Some Key Issues in the Philosophy of Science*, University of Chicago Press, Chicago.
- LEEDS, S.** 1978, "Theories of Reference and Truth", *Erkenntnis*, **13**, pp. 111-129.
- LePORE, E.** 1983, "What Model Theoretic Semantics Cannot Do", *Synthese*, **54**, pp. 167-188.
- LINDSTRÖM, S.** 2001, "Horwich's Minimalist Conception of Truth: Some Logical Difficulties", *Logic and Logical Philosophy*, **9**, pp. 161-181.
- LYNCH, M.P.** (ed.) 2001, *The Nature of Truth: Classic and Contemporary Perspectives*, MIT Press, Cambridge, Massachusetts.
- MACDONALD, G. & WRIGHT, C.** (eds.) 1986, *Fact, Science and Morality: Essays on A.J. Ayer's Language, Truth and Logic*, Blackwell, Oxford.
- MACDONALD, M.** (ed.) 1954, *Philosophy and Analysis*, Blackwell, Oxford.
- MacKENZIE, P.T.** 2005, "Truth and the Magic of 'Is'", *Philosophy*, **80**, pp. 125-134.
- MADDY, P.** 1988, "Believing the axioms, I", *Journal of Symbolic Logic*, **53**, pp. 481-511.
- MARTIN, R.L.** (ed.) 1984, *Recent Essays on Truth and the Liar Paradox*, Oxford University Press, Oxford.
- McGEE, V.** 1992, "Maximal Consistent Sets of Instances of Tarski's Schema (T)", *Journal of Philosophical Logic*, **21**, pp. 235-241.
- MILNE, P.** 1999, "Tarski, Truth and Model Theory", *Proceedings of the Aristotelian Society*, **XCIX**, pp. 141-167.
- MONK, J.D.** 1976, *Mathematical Logic*, Springer-Verlag, New York.

**MONTAGUE, R.** 1963, “Syntactic Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability” , *Acta Philosophica Fennica*, **16**, pp. 153-167.

**MOORE, G.E.**

- 1908, “William James’ ‘Pragmatism’”, *Proceedings of the Aristotelian Society*, **8**.  
Reprinted in his *Philosophical Studies*, pp. 97-146.
- 1953, *Some Main Problems in Philosophy*, Allen & Unwin Ltd., London.
- 1959, *Philosophical Studies*, Littlefield, Adams & Co., Paterson, New Jersey.

**MOSTOWSKI, A.** 1966, *Thirty Years of Foundational Studies: Lectures on the Development of Mathematical Logic and the Study of the Foundations of Mathematics in 1930-1964*, Blackwell, Oxford.

**MOU, B.** 2000, “A Metaphilosophical Analysis of the Core Idea of Deflationism”, *Metaphilosophy*, **31**, pp. 262-286.

**MUSGRAVE, A.** 1997, “The T-Scheme Plus Epistemic Truth Equals Idealism”, *Australasian Journal of Philosophy*, **75**, pp. 490-496. Reprinted in Salamun, K. (ed.) 1999, pp. 185-192.

**NAGEL, T.** 1974, “What Is It Like to Be a Bat?”, *Philosophical Review*, **83**, pp. 435-450.

**NEALE, S.**

- 1995, “The Philosophical Significance of Gödel’s Slingshot”, *Mind*, **104**, pp. 761-825.
- 1997 (with J. Dever), “Slingshots and Boomerangs”, *Mind*, **106**, pp. 143-168.
- 2001, *Facing Facts*, Clarendon Press, Oxford.

**NEURATH, O.**

- 1931, “Physikalismus”. Translated as “Physicalism” in Neurath, O. 1983, pp. 52-57.
- 1932, “Protokollsätze”, *Erkenntnis*, **3**, pp. 204-214. Translated as “Protocol Sentences” in Neurath, O. 1983, pp. 91-99.
- 1934, “Radikaler Physikalismus und ‘Wirklicher Welt’ ”, *Erkenntnis*, **4**, pp. 346-362.  
Translated as “Radical Physicalism and the ‘Real World’ ” in Neurath, O. 1983, pp. 100-114.
- 1983, *Philosophical Papers: 1913-1946*, edited by R. Cohen & M. Neurath, Reidel, Dordrecht.

**NIINILUOTO, I.**

- 1987, *Truthlikeness*, Reidel, Dordrecht.
- 1999a, "Tarskian Truth as Correspondence - Replies to Some Objections", in Peregrin, J. (ed.) 1999, pp. 91-104.
- 1999b, "Theories of Truth: Vienna, Berlin and Warsaw" in Woleński, J & Köhler, E. (eds.) 1999, pp. 17-26.
- 2002, *Critical Scientific Realism*, Oxford University Press, Oxford.
- 2004, "Tarski's Definition and Truth-Makers", *Annals of Pure and Applied Logic*, **126**, pp. 57-76.

**O'LEARY-HAWTHORNE, J.** (with H. Price) 1996, "How to Stand Up for Non-Cognitivists", *Australasian Journal of Philosophy*, **74**, pp. 275-292.

**OPPY, G.**

- 1997, "The Philosophical Insignificance of Gödel's Slingshot", *Mind*, **106**, pp. 121-142.
- 2004, "Facing Facts?", *Australasian Journal of Philosophy*, **82**, pp. 621-643.

**ORWELL, G.** 1949, *Nineteen Eighty-Four*, Penguin, London.

**PATTERSON, D.** 2003, "What is a Correspondence Theory of Truth", *Synthese*, **137**, pp. 421-444.

**PEARCE, D.** 1987, *Roads to Commensurability*, Reidel, Dordrecht.

**PEIRCE, C.S.**

- 1878, "How to Make our Ideas Clear", *Popular Science Monthly*, **12**, pp. 286-302.  
Reprinted in Lynch, M.P. (ed.) 2001, pp. 193-209 and in Schmitt, F.F. 2004a, pp. 41-58.
- 1905, "What Pragmatism Is" in Hartshorne, C. & Weiss, P. (eds.) 1960, *Collected Papers of Charles Sanders Peirce, Volumes 1-8*.
- 1960, *Collected Papers of Charles Sanders Peirce, Volumes 1-8*, edited by C. Hartshorne and P. Weiss, Harvard University Press, Cambridge, Massachusetts.
- 1966, "Letters to Lady Welby" in P.P. Wiener (ed.) 1966, *Charles S. Peirce: Selected Writings*, Dover, New York.

**PENDLEBURY, M.** 1986, "Facts as Truthmakers", *Monist*, **69**, pp. 177-188.

**PEREGRIN, J.** (ed.) 1999, *Truth and its Nature (if any)*, Kluwer, Dordrecht.

**PITCHER, G.** (ed.) 1964, *Truth*, Prentice-Hall, Englewood Cliffs.

**PLOTINUS** c.250, *Enneades I – III*. In *Opera*, Clarendon Press, Oxford 1964.

**POPPER, K.R.**

- 1959, *The Logic of Scientific Discovery*, Routledge, London
- 1960, *Conjectures and Refutations*, Routledge, London. Revised edition 1972.
- 1972, *Objective Knowledge: An Evolutionary Approach*, Clarendon Press, Oxford. Revised edition (reprinted with corrections and a new appendix 2) 1979.
- 1974a, "Some Philosophical Comments on Tarski's Theory of Truth", *Proceedings of the Tarski Symposium*, pp. 400-402, American Mathematical Society, Providence.
- 1974b, "Intellectual Autobiography", opening chapter of P.A. Schilpp (ed.) 1974, *The Philosophy of Karl Popper*, Open Court, La Salle, Illinois.

**PRIOR, A.** 1971, *Objects of Thought*, Clarendon Press, Oxford.

**PRZEŁECKI, M.** 1969, *The Logic of Empirical Theories*, Routledge & Kegan Paul, London.

**PUTNAM, H.**

- 1975a, *Mathematics, Matter and Method: Philosophical Papers (Volume 1)*, Cambridge University Press, Cambridge. Second edition, 1979.
- 1975b, *Mind, Language and Reality: Philosophical Papers (Volume 2)*, Cambridge University Press, Cambridge. Second edition, 1979.
- 1981, *Reason, Truth and History*, Cambridge University Press, Cambridge.

**QUINE, W.V.O.**

- 1941, "Whitehead and the Rise of Modern Logic", in Quine 1995b, pp. 3-36.
- 1953a, "Reference and Modality", in Quine 1980, pp. 139-159.
- 1953b, "Three Grades of Modal Involvement", in Quine 1966, pp. 158-176.
- 1960, *Word and Object*, MIT Press, Cambridge, Massachusetts.
- 1966, *The Ways of Paradox*, Harvard University Press, Cambridge, Massachusetts.

- 1970, *Philosophy of Logic*, Prentice Hall, Englewood Cliffs, New Jersey. Second edition published by Harvard University Press, Cambridge, Massachusetts in 1986.
- 1972, *Methods of Logic*, Harvard University Press, Cambridge, Massachusetts, Third Edition.
- 1974a, *The Roots of Reference*, Open Court, La Salle, Illinois.
- 1974b, "Truth and Disquotation", *Proceedings of Symposia in Pure Mathematics*, **25**, pp. 373-384. An amended version is reprinted in Quine, W.V.O. 1976, pp. 308-321.
- 1976, *The Ways of Paradox*. Revised enlarged second edition. Harvard University Press, Cambridge, Massachusetts.
- 1980, *From a Logical Point of View: Nine-Logico-Philosophical Essays*. Revised second edition. Harvard University Press, Cambridge, Massachusetts.
- 1981, *Theories and Things*, Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- 1987, *Quiddities: An Intermittently Philosophical Dictionary*, Penguin, London.
- 1992, *Pursuit of Truth*. Revised edition. Harvard University Press, Cambridge, Massachusetts. First edition 1990.
- 1995, *From Stimulus to Science*, Harvard University Press, Cambridge, Massachusetts.

#### **RAATIKAINEN, P.**

- 2005, "On Horwich's Way Out", *Analysis*, **65**, pp. 175-177.
- (forthcoming, i), "Problems of Deflationism" in T. Aho et al (eds.), *Truth and Games in Logic and Language*.
- (forthcoming, ii), "Truth, Correspondence, Models and Tarski"

#### **RAMSEY, F.P.**

- 1923, 'Critical Notice of *Tractatus Logico-Philosophicus*' in Ramsey, F.P. 1931, pp. 270-286.
- 1927, "Facts and Propositions", *Proceedings of the Aristotelian Society*, supplementary volume, **7**, pp. 153-170. Reprinted in Pitcher, G. (ed.) 1964, pp. 16-17.
- 1931, *The Foundation of Mathematics and Other Essays*, edited by R.B.Braithwaite, Routledge, London.
- 1990, "The Nature of Truth", *Episteme*, **16**, pp. 6-16. Reprinted in Lynch, M.P. (ed.) 2001, pp. 433-446.

**READ, S.** 1995, *Thinking about Logic: An Introduction to the Philosophy of Logic*, Oxford University Press, Oxford.

**RECK, E.H.** (ed.) 2002, *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, Oxford University Press, Oxford.

**REID, C.** 1970, *Hilbert*, Springer, Berlin.

**RESTALL, G.** 2004, "One Way to Face Facts", *Philosophical Quarterly*, **54**, pp. 420-426.

**RODRIGUEZ-PEREYRA, G.** 2003, "Review of S. Neale 2001, *Facing Facts*", *Mind*, **112**, pp. 780-786.

**ROJSZCZAK, A.**

- 1999, "Why Should a Physical Object Take on the Rôle of Truth-Bearer?" in Woleński, J. & Köhler, E. (eds.) 1999a, 115-125.
- 2002, "Philosophical Background and Philosophical Content of the Semantic Definition of Truth", *Erkenntnis*, **56**, pp. 29-62.

**RORTY, R.** 1986, "Pragmatism, Davidson and Truth" in E. Lepore (ed.) 1986, *Truth and Interpretation*, Blackwell, Oxford. Reprinted in Blackburn & Simmons (eds.) 1999, pp. 323-350.

**RYLE, G.**

- 1949a, *The Concept of Mind*, Hutchinson, London.
- 1949b, "Meaning and Necessity", *Philosophy*, **24**, pp. 69-76.

**RUSSELL, B.A.W.**

- 1905, "On Denoting", *Mind*, **14**, pp. 479-493. Reprinted in Russell, B.A.W. 1956, pp. 41-56.
- 1910a, *Philosophical Essays*, George Allen & Unwin, London. Latest edition 1966.
- 1910b, "William James's Conception of Truth" in Russell, B.A.W. 1910a (1966), pp. 112-130. Reprinted in Blackburn & Simmons (eds.) 1999, pp. 69-82.
- 1910c, "The Monistic Theory of Truth" in Russell, B.A.W. 1910a (1966), pp. 131-146.
- 1910d, "On the Nature of Truth and Falsehood" in Russell, B.A.W. 1910a (1966), pp. 147-159.

- 1912a, *The Problems of Philosophy*, Williams and Norgate, London. Latest edition 1967, Oxford University Press, Oxford.
- 1912b, "On the Notion of Cause", *Proceedings of the Aristotelian Society*, **13**, pp. 1-26. Reprinted as Chapter 10 in his *Mysticism and Logic*, 10<sup>th</sup> impression, Allen & Unwin, London, 1986.
- 1918, "The Philosophy of Logical Atomism", in Russell, B.A.W. 1956, pp. 177-281.
- 1940, *An Inquiry into Meaning and Truth*, Allen & Unwin Ltd., London.
- 1948, *Human Knowledge: Its Scope and its Limits*, Allen & Unwin, London.
- 1956, *Logic and Knowledge: Essays 1901-1950*. Allen & Unwin Ltd., London, 1988.
- 1959, *My Philosophical Development*, Routledge, London.

**SALAMUN, K.** (ed.) 1999, *Alan Musgrave: Essays on Realism and Rationalism*, (Volume XII of the series in the Philosophy of Karl R. Popper and Critical Rationalism), Rodopi, B.V., Amsterdam-Atlanta, GA.

**SCHANTZ, R.** 2002, (ed.) *Current Issues in Theoretical Philosophy: What is Truth?*, Volume 1, Walter de Gruyter, Berlin.

**SCHEFFLER, I.** 1967, *Science and Subjectivity*, Bobbs-Merrill Co., Indianapolis.

**SCHILLER, F.C.S.** 1907, *Studies in Humanism*, MacMillan & Co., London.

**SCHLICK, M.**

- 1934, "Über das Fundament der Erkenntnis", *Erkenntnis*, **4**, pp. 79-99.
- 1935, "Facts and Propositions", *Analysis*, **2**, pp. 65-70.
- 1979, *Philosophical Papers*, Reidel, Dordrecht.

**SCMITT, F.F.**

- 1995, *Truth: A Primer*, Westview Press, Boulder.
- 2004, *Theories of Truth*, Blackwell, Oxford.

**SEARLE, J.R.** 1995, *The Construction of Social Reality*, Penguin, London.

**SELLARS, W.** 1997, *Empiricism and the Philosophy of Mind*, Harvard University Press, Cambridge, Massachusetts.

**SHAPIRO, S.**

- 1983, "Conservativeness and Incompleteness", *Journal of Philosophy*, **80**, pp. 521-531. Reprinted in Hart, W. (ed.) 1996, pp. 225-234.
- 1998, "Proof and Truth- Through Thick and Thin", *Journal of Philosophy*, **95**, pp. 493-522.
- 2002, "Deflation and Conservation", in V. Halbach, & L. Horsten, (eds.), *Principles of Truth*, pp. 102-128.

**SHER, G.**

- 1999, "On the Possibility of a Substantive Theory of Truth", *Synthese*, **117**, pp. 133-172.
- 2001, "Truth, Logical Structure and Compositionality", *Synthese*, **126**, pp. 195-219..

**SHEARD, M.** 1994, "A Guide to Truth Predicates in the Modern Era", *Journal of Symbolic Logic*, **59**, pp. 1032-1054.

**SLUGA, H.** 2002, "Frege on the Indefinability of Truth" in Reck, E.H. (ed.) 2002, pp. 75-95.

**SMULLYAN, R.** 1948, "Reference and Modality", *Journal of Symbolic Logic*, **13**, pp. 31-37.

**SOAMES, S.**

- 1984, "What is a Theory of Truth?", *Journal of Philosophy*, **81**, pp. 411-429. Reprinted in Lynch, M.P. (ed.) 2001, pp. 397-418.
- 1997, "The Truth about Deflationism" in E. Villanueva (ed.), *Philosophical Issues*, **viii**, pp. 1-44. Ridgeview Press, Atascadero.
- 1999, *Understanding Truth*, Oxford University Press, Oxford.

**STOLTJAR, D. & DAMNJANOVIC, N.** 2007, "The Deflationary Theory of Truth", entry in *The Stanford Encyclopedia of Philosophy* (edited by E.N. Zalta).

**STRAWSON, P.F.**

- 1949, "Truth", *Analysis*, **9**, pp. 82-97. Reprinted in M. Macdonald (ed.) 1954, *Philosophy and Analysis*, pp. 260-277, Blackwell, Oxford.
- 1950, "Truth", *Proceedings of the Aristotelian Society*, supplementary volume, **24**, pp. 129-156. Reprinted in Pitcher, G. (ed.) 1964, pp. 32-53, also in Strawson, P. 1971, pp. 190-213 and (with postscript) in Blackburn, S. & Simmons, K. (eds.) 1999, pp. 162-182.



- 1964, "A Problem About Truth - A Reply To Mr. Warnock", in Pitcher, G. (ed.) 1964, pp. 68-84. Reprinted in Strawson, P. 1971, pp. 214-233.
- 1971, *Logico-Linguistic Papers*, Methuen, London.
- (ed.) 1973, *Philosophical Logic*, reprinted edition (first edition 1967), Oxford University Press, Oxford.

**SUNDHOLM, G.** 1994, "Ontologic versus Epistemologic: Some Strands in the Development of Logic, 1837-1957" in D. Prawitz & D. Westerståhl (eds.) 1994, *Logic and Philosophy of Science in Uppsala*, Kluwer, Dordrecht, pp. 373-384.

**SZANIAWSKI, K.** (ed.) 1989, *The Vienna Circle and the Lvov-Warsaw School*, Kluwer, Dordrecht.

**SZUBKA, T.** 2003, "The T-Schema and the Epistemic Conception of Truth", *Teorema*, **22**, pp. 93-99.

**TARSKI, A.**

- 1934, "Some Methodological Investigations on the Definability of Concepts", in Tarski, A 1983, pp. 296-319.
- 1936a, "Der Wahrheitsbegriff in den Formalisierten Sprachen", *Philosophia Studia*, **1**, pp. 261-405. The English translation "The Concept of Truth in Formalized Languages" is reprinted in Tarski, A 1983, pp. 152-278.
- 1936b, "O ugruntowaniu naukowej semantyki", *Przegląd Filozoficzny*, **39**, pp. 50-57. Translated into English as "The Establishment of Scientific Semantics" by J.H. Woodger and reprinted in Tarski, A. 1983, pp. 401-408.
- 1944, "The Semantic Conception of Truth and The Foundations of Semantics", *Philosophy and Phenomenological Research*, **4**, 105-112. Reprinted in Blackburn, S. & Simmons, K. (eds.) 1999, pp. 115-143. Also reprinted in Lynch, M.P. (ed.) 2001, pp. 331-363 and in Schmitt, F.F. 2004a, pp. 115-151.
- 1954-55, "Contributions to the Theory of Models I-III", *Indagationes Mathematicae*, **16**, pp.572-581, 582-688; **17**, pp. 56-64.
- 1957 (with R. Vaught), "Arithmetical Extensions of Relational Systems", *Compositio Mathematicae*, **13**, pp. 81-102.
- 1969, "Truth and Proof", *Scientific American*, **220**, pp. 63-77.

- 1983, *Logic, Semantics, Metamathematics: Papers by Alfred Tarski from 1922-1938*, edited by J. Corcoran, translated by J.H. Woodger. Hackett Publishing, Indianapolis. First edition 1956, Clarendon Press, Oxford.

**TENNANT, N.**

- 2002, "Deflationism and the Gödel Phenomena", *Mind*, **111**, pp. 551-582.
- 2005, "Deflationism and the Gödel Phenomena: Reply to Ketland", *Mind*, **114**, pp. 89-96.

**TUOMELA, R.** 1973, *Theoretical Concepts*, Springer-Verlag, Berlin.

**VAUGHT, R.** 1974, "Model Theory before 1945" in Henkin et al (eds.) 1974, *Proceedings of the Tarski Symposium*, American Mathematical Society, Providence, RI, pp. 153-172.

**VEBLEN, O.** 1904, "A System of Axioms for Geometry", *Transactions of the American Mathematical Society*, **V**, pp. 343-381.

**VISION, G.** 2004, *Veritas: The Correspondence Theory and Its Critics*, Massachusetts Institute of Technology, Cambridge, Massachusetts.

**WALKER, R.C.S.** 1999, "Theories of Truth", in Hale, B. & Wright, C. (eds.) 1999, pp. 309-330.

**WARNOCK, G.J.** 1964, "A Problem About Truth", in Pitcher, G. (ed.) 1964, pp. 54-67.

**WHITE, A.R.** 1970, *Truth*, Routledge, London.

**WHITEHEAD, A.N.** 1925 (with B. Russell), *Principia Mathematica (volume i)*, Second Edition, Cambridge University Press, Cambridge.

**WILLIAMS, C.J.F.** 1976, *What is Truth?*, Cambridge University Press, Cambridge.

**WILLIAMS, M.**

- 1986, "Do We (Epistemologists) Need a Theory of Truth?", *Philosophical Topics*, **14**, pp. 223-242.
- 1988, "Epistemological Realism and the Basis of Skepticism", *Mind*, **85**, pp. 415-439.
- 1999, "Meaning and Deflationary Truth", *Journal of Philosophy*, **96**, pp. 545-564.

- 2002, "On Some Critics of Deflationism" in Schantz, R. (ed.) 2002, pp. 146-158.

**WILSON, W.K.** 1990, "Some Reflections on the Prosentential Theory of Truth", in Dunn, J.N. & Gupta, A. (eds.) 1990, pp. 19-32.

**WITTGENSTEIN, L.J.J.**

- 1922, *Tractatus Logico-Philosophicus*, translated by C.K. Ogden. Paperback edition 1981, Routledge, London.
- 1958, *Philosophische Untersuchungen/Philosophical Investigations*, translated by G.E.M. Anscombe, Blackwell, Oxford. Reprinted third edition 1967.

**WOZLEY, A.D.** 1966, *Theory of Knowledge*, Barnes & Noble, London.

**WOLEŃSKI, J.**

- 1989a (with P. Simons), "De Veritate: Austro-Polish Contributions to the Theory of Truth from Brentano to Tarski" Szaniawski, K. (ed.) 1989, pp. 391-442.
- 1989b, "The Lvov-Warsaw School and the Vienna Circle" in Szaniawski, K. (ed.) 1989, pp. 443-453.
- 1993, "Tarski as a Philosopher" in Coniglione, F., Poli, R. & Woleński, J. (ed.s) 1993, *Polish Scientific Philosophy: The Lvov-Warsaw School*, Rodopi, Amsterdam, pp. 319-338.
- 1998 (editor with K. Kijania-Placek), *The Lvov-Warsaw School and Contemporary Philosophy*, Kluwer, Dordrecht.
- 1999 (editor with E. Köhler), *Alfred Tarski and the Vienna Circle: Austro-Polish Connections in Logical Empiricism, Papers from the International Conference held in Vienna, July 12<sup>th</sup>-14<sup>th</sup>, 1997*, Kluwer, Dordrecht.
- (web article), "Theories of Truth in Austrian and Polish Philosophy", <http://www.fmag.unict.it/~polphil/PolPhil/LvovWarsaw/WolTruth.html>.

**WRIGHT, C.W.**

- 1992, *Truth and Objectivity*, Harvard University Press, Cambridge, Massachusetts.
- 1999, "Truth: A Traditional Debate Reviewed", in Blackburn, S. & Simmons, K. (eds.) 1999, pp. 203-238.