Tarski's Threat to the T-Schema

MSc Thesis (*Afstudeerscriptie*)

written by

Cian Chartier
(born November 10th, 1986 in Montréal, Canada)

under the supervision of **Prof. dr. Frank Veltman**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:
August 30, 2011

Members of the Thesis Committee:
Prof. Dr. Frank Veltman
Prof. Dr. Michiel van Lambalgen
Prof. Dr. Benedikt Löwe
Dr. Robert van Rooij

Institute for Logic, Language and Computation

Abstract

This thesis examines truth theories. First, four relevant programs in philosophy are considered. Second, four truth theories are compared according to a range of criteria. The truth theories are categorised according to Leitgeb's eight criteria for a truth theory. The four truth theories are then compared with each-other based on three new criteria. In this way their relative usefulness in pursuing some of the aims of the four programs is evaluated. This presents a springboard for future work on truth: proposing ideas for different truth theories that advance unambiguously different programs on the basis of the four truth theories proposed.

Acknowledgements

I first thank my supervisor Prof. Dr. Frank Veltman for regularly reading and providing constructive feedback, recommending some papers along the way. This thesis has been much improved by his corrections and input, including extensive discussions and suggestions of the ideas proposed here. His patience in working with me on the thesis has been much appreciated. Among the other professors in the ILLC, Dr. Robert van Rooij also provided helpful discussion. My academic tutor Dr. Maria Aloni and Dr. Alessandra Palmigiano gave useful advice to me while I was considering my topic.

I attended the Truth Be Told conference in Amsterdam (March 23-25, 2011), which provided several insights into cutting-edge research on truth. I am indebted to Theodora Achourioti and Peter van Ormondt for organising the conference. Several of the speakers have work cited here, though special acknowledgement goes to Prof. Hartry Field for clarifying a few points.

There is a good intellectual atmosphere in Science Park and the Master of Logic room in particular. Many of my fellow Master of Logic students provided helpful comments of support. My memory is far from exhaustive, but I can recall a few particularly challenging discussions with Matthew Wrampler-Doty, Alwin Blok, and Riccardo Pinosio on the ideas leading up to the creation of this thesis.

I am thankful to Tanja Kassenaar for her calmness and efficiency when making fairly late preparations for my thesis defense. I especially appreciate her ability to maintain a serene disposition even when faced with questions already answered on the ILLC web site.

Last but not least are the many people from outside of logic who I tried to explain my thesis to. Chief among these are my mother Prof. Eithne Guilfoyle and Alaina Piro Schempp who read my thesis in an almost-complete state.

Many thanks go to anyone else I may have forgotten who played their own role in the eventual creation of this thesis.

Contents

1 Introduction

1.1 Background and aims

The problems that arise from defining a truth predicate T acting on a *truth-bearer* ϕ, such that the formula $T(\ulcorner phi \urcorner) \equiv \phi$ is satisfied - the formula we refer to as an instance of the *T-schema* - are not only central in the programs of natural language semantics of Russell and Davidson, but also play an important rule in the philosophical debate on deflationary theories of truth. Among these problems, we focus on results from Tarski, particularly *Tarski's undefinability theorem*, that show that such a predicate cannot co-exist with the axioms and inferences of classical logic on pain of triviality. Tarski's own conclusion was that the *ordinary notion* of truth was incoherent, and should be replaced with the coherent notion denoted by *typed* predicates that he went on to provide. Kripke called into question the typed predicates' applicability in either natural language semantics or philosophy, and wrote the influential [47] with a single *type-free* truth predicate. Since then many logics, or more specificially *truth theories* with type-free truth predicates, have been advanced.

This thesis shall examine truth theories. First, four relevant programs in philosophy will be considered. Second, four truth theories will be compared according to a range of criteria. The truth theories will be categorised according to Leitgeb's eight criteria for a truth theory. The four truth theories will then be compared with each-other based on three new criteria. In this way their relative usefulness in pursuing some of the aims of the four programs will be evaluated. This shall present a springboard for future work on truth: proposing ideas for different truth theories that advance unambiguously different programs on the basis of the four truth theories proposed.

This thesis adopts a "bird's eye view" approach in evaluating the relative merits of several truth theories on the basis of the range of criteria. This is in contrast to much work on truth, which tends to advance a single stance or truth theory against all others. One advantage of the bird's eye view approach is that it allows us to categorise truth theories based on their relevance in solving a particular problem. There is not just one problem motivating the need for a truth theory but, as we shall see, several, and different problems may require different solutions.

Another advantage of the bird's eye view approach is it allows us to

clear up a presently convoluted discussion of the comparison of truth theories. Indeed many books and papers have their own substantially different notions of what it means for a formula or natural language sentence to be "grounded" or "paradoxical". I shall be as careful as I can to define each term unambiguously, for the sake of this discussion. Anyone with training in first-order logic (see [23]), model theory (see the early chapters of [22]), coding of formulas (see [20]), and models of set theory (see the relevant chapters of [21]) should hopefully be capable of understanding the thesis, though prior familiarity with Tarski's work in [48] and Kripke's work in [47] is strongly recommended.

1.2 Structure of the thesis

In this second section of the thesis, I shall motivate why people should be interested in truth theories: first why we should be interested in the T-schema, what damaging role semantic paradoxes have to play, and why we should find notions of "groundedness" and "contingently paradoxical formulas/sentences" compelling, leading us to the truth theories mentioned in this thesis. There shall also be a note on other perceived threats to the T-schema when the truth-bearers are natural language sentences.

In the third section of the thesis, I shall reflect on the difficulties posed by the theorems of Tarski and McGee by investigating the theoretical options left for truth theories to pursue. The options shall be set out as three categories of truth theory based on which of Leitgeb's (unsatisfiable) list of eight criteria in [14] can be met. Having set out our categories, I shall set out three criteria to evaluate the truth theories in each category. The "ghost" of Tarski's theorem shall be seen again in the form of *revenge paradoxes*.

In the fourth section, a case study shall be made of Field's theory of truth. Here the truth theory is not classical, and satisfies the T-schema with respect to biconditionals - not material biconditionals, but rather those of Field's own creation. The truth predicate is not fully compositional with respect to the conditional operator we consider here, but there is current research going on to replace the conditional with one that respects compositionality. The classical law that is violated is the *law of excluded middle*, with sentences diagnosed as paradoxical with respect to a given situation being the exceptions to the law; as such it is a *paracomplete* theory. Field's theory satisfies particularly strong intersubstitutivity principles with respect to the

truth predicate, giving it an advantage in certain respects over other theories under discussion. However we shall call into question whether Field's logic (or indeed any paracomplete theories) truly is "revenge-immune", and the cognitive plausibility of Field's new conditional operator.

In the fifth section, Priest's logic of paradox will be investigated. The logic of paradox is *paraconsistent*, and as such contradictions can be derived from the liar sentence, but without the disjunctive syllogism (modus ponens for the material implication) from which a contradiction can be derived. The logic of paradox is the dual of the Strong Kleene three-valued logic K_3, and the truth predicate can be defined via a dual process to that of Kripke's fixed points. The T-schema biconditionals are always declared true in the logic of paradox, though sometimes they are also declared false. We shall see that without (at least) either a departure into a logic with an infinite number of truth values or a radical revision of the metatheoretic apparatus, the logic of paradox is still open to revenge paradoxes. The apparent duality of many paraconsistent theories with paracomplete ones will also be discussed, calling into question the necessity of rejecting the law of non-contradiction as a matter of principle.

In the sixth section, the set-theoretic truth theory of Leitgeb and Welch will be the central case study. This theory is an instance of a classical truth theory with a compositional truth predicate, where certain instances of theorems of the truth theory are not declared true. Leitgeb and Welch have a logic of propositions from which ZF set theory can be derived, and also a principled way of determining whether a sentence in a language expresses a proposition on the basis of whether it is semantically *grounded* in a sense they define independently of Kripke. The intention of their truth theory is to provide a foundation for semantics. The truth theory is in some sense revenge-immune at the cost of limiting the truth-bearers to grounded sentences. It shall be argued that groundedness is too strong a constraint on a sentence for it to be meaningful or even truthful according to a sensible foundation of semantics.

In the seventh section, a few different revision theories of truth will be distinguished and given a general analysis. Each of these satisfy classical rules, but in one respect or another the respective truth predicate is not compositional.[1] One reading of the revision theories respects the un-

[1]The Friedman-Sheard revision theory in [4] has a compositional truth predicate, but

restricted T-schema, but only with respect to definitional equivalence with respect to a theory of circular definitions. The T-schema with respect to material biconditionals is necessarily restricted in the revision theory. Some relatively convoluted revenge paradoxes can also be developed for at least some revision theories by including "determinacy" operator in the language. And just how applicable or intuitive each of the theories are, and also how to choose between seemingly arbitrary choices of "limit rule"s that determine the scope of the truth predicates, will be called into question.

In the final section, conclusions will be drawn in two respects. First, with regards to which theories are more promising for solving which problems. The proposed paracomplete and paraconsistent theories will be construed as two sides of the same coin, and more able to support truth as a device of generalisation through a principle of intersubstitutivity. Second, in seeing how theoretical tools of certain truth theories can help others towards solving the problems which they have been assigned. The Leitgeb-Welch theory's constraints on truth-bearing sentences will be relaxed and include a broadly revision-theoretic apparatus to declare whether a sentence is, if not expressing a proposition, at least "stable" (to distinguish liars from truth-tellers). This we will argue to have some promising beginnings for a foundation of semantics, as Leitgeb and Welch intended.

The overall conclusion of the thesis is, briefly, a tentative suggestion that certain truth theories may suit as an adequate springboard for particular philosophical programs rather than just a study of the paradoxes in and of themselves. The four programs mentioned in the second section shall be coupled with three of the four truth theories covered here. A speculative table grouping the programs with the suitable truth theories appears below:

Programs	Suitable theories
Foundation of semantics + Philosophical logic	Leitgeb-Welch
Deflationary stance + Naive property theory	Field/Priest

at the expense of ω-consistency and thus condition C4, so it will not be discussed here.

2 Motivation

2.1 Why do we want a truth predicate?

The subject of this thesis will be logics of *truth theories* in the following sense: a *truth predicate* is a predicate T (or "true") in a language L such that for any L-sentence ϕ without any instances of T, $T(\ulcorner\phi\urcorner) \equiv \phi$, and a *truth theory* is a deductive L-theory. We consider "$T(\ulcorner\phi\urcorner) \equiv \phi$ for ϕ an L-sentence without any instances of T" to be a restricted form of the *T-schema*, and can be thought of as a minimal constraint on truth predicates; if there are no added restrictions on L-sentences ϕ, then the T-schema is said to be *unrestricted*. For most of the thesis L will be considered to be a first-order language (so ϕ is a first-order sentence), though in some of this section L may be thought of as containing higher-order quantifiers or even informally as a natural language such as English. We shall distinguish between sentences of some logical language L and *natural language sentences*.

The basis of this subsection is to answer why we should be interested in logical languages with truth predicates. To do this, we briefly explain at least four different areas of study with respect to which the question has some relevance:

A1 To provide the meaning of any of its sentences from a finite set of axioms (based on the conditions on which those sentences are true).

A2 To have a logical language which can express every coherent notion expressed in natural language.

A3 To strengthen (or replace) ZF set theory with a naive property theory, having resolved the paradoxes within.

A4 Providing support for the deflationary stance on truth.

A1 is the program of *Davidsonian semantics* in [42], aiming to obtain a theory of meaning from a theory of truth. A2 can be seen as part of the "philosophical logic" program of Bertrand Russell. A3 implicitly holds with it the promise of a foundation of mathematics and semantics treating Russell's paradox in a similar fashion as it would treat the liar paradox. One of the later theories investigated in this section does both in the exact same fashion. Most of the work on developing A3 is beyond the scope of this thesis, but it will be mentioned again at the end of this thesis. A breakthrough in this area may give rise to an alternative foundation of mathematics to set

theory.

A4 is the one point where this thesis touches discussions of *philosophical theories of truth* providing a metaphysical account of what truth is, e.g. deflationary theories, inflationary theories, correspondence, and so on. The *deflationary theory of truth* is construed here as family of theories that postulate that for any sentence ϕ in any language L with a truth predicate T, $T(\ulcorner\phi\urcorner) \equiv \phi$, in an unrestricted form, exhaustively defines truth. This is Quine's *disquotational schema*. Thus to assert that a sentence is true is nothing more than to assert the sentence itself. Numerous truth theories accord with a deflationary theory's account according to one reading or another of the equivalence symbol $\equiv$. For example, Anil Gupta's revision theory of truth from [36] is a truth theory in accord with the deflationary theory of truth, but only if the T-schema equivalences are read the same way as definitional identity $=$ in Gupta's theory of circular definitions - not if they are read the same way as the material bi-implication $\leftrightarrow$.

It may seem peculiar that a truth predicate's existence has to be justified in order for truth to be seen as "redundant" in Ayer's sense. But the existence of the semantic paradoxes discussed in 2.2 nonetheless poses a threat to the coherence of the deflationary stance on truth, as argued in [43]. And at any rate, deflationists typically agree that the truth predicate T is useful in logical languages for purposes of generalisation, which will be mentioned in 2.3.

One area not mentioned is that of studies of vagueness, and the Sorites Paradox in particular. Several "theories of truth" also work as "theories of vagueness" with often only very slight modifications. Some work suggests that breakthroughs in the study of truth might inform the study of vagueness, but these connections go outside the scope of this thesis; examples abound in [52].

Now that we've seen how the study of a truth predicate can be fruitful, we shall move on to two theorems of Tarski and McGee on formal languages L that impose limitations on what truth theories can include.

2.2 The paradoxes; negative results of Tarski and McGee

Here we provide details of two important results that impose limitations on the properties that a truth predicate in a first-order language can have. These make the task of having a theory of truth for first-order sentences, never mind natural language sentences, a challenging and obscure one at the outset. But in order to understand the nature of the challenge, some definitions and conditions should first be set.

Here we generally work in a language L with a single truth predicate T that acts on coded names of formulas[2]. Truth is said to be *type-free* if it is represented by a single truth predicate; if there is an (integer-valued) hierarchy of truth predicates T_α to be introduced such that T_0 acts on all coded names of formulas without any T_α, and $T_{\alpha+1}$ acts on all coded names of formulas without any T_β for $\beta \geq \alpha$, then truth is said to be *typed*. Let L' be a language with a typed hierarchy of truth predicates. Type-free truth theories are theories in L, and typed truth theories are first-order theories in L'.

When we speak of the *outer logic*, we are referring to the logic of the truth theory. In contrast when we speak of the *inner logic*, we are referring to the logic of the set of sentences (the *inner theory*) on whose coded names the truth predicate or predicates of the truth theory act. It is conceivable, though peculiar, that the inner logic and the outer logic may not be the same (e.g. the law of excluded middle may be derivable in the theory of truth, but it may not be considered true) so it is important to distinguish the two.

Throughout much of the thesis, we shall assume that the truth theory and its inner theory contain the axioms of Peano Arithmetic. It is necessary for a good truth theory to be consistently added to the theory of Peano Arithmetic, yet this assumption is already sufficient to lead us into trouble:

Theorem (Tarski): Any type-free classical truth theory that contains the theory of Peano Arithmetic and the unrestricted T-schema, with $\equiv$ meaning two-way material implication $\leftrightarrow$, is trivial.

Proof: By Gödel's Diagonal Lemma, for every formula $\phi(x)$ there is a sentence ψ such that $\psi \leftrightarrow \phi(\ulcorner\psi\urcorner)$ is contained in the truth theory. Let $\phi(x)$ be $\neg T(x)$; then there is a sentence ψ such that $\psi \leftrightarrow \neg T(\ulcorner\psi\urcorner)$ is contained in the truth theory. Assuming the law of excluded middle, we have

[2]The coding scheme in [20] would be sufficient

$T(\ulcorner\psi\urcorner) \vee \neg T(\ulcorner\psi\urcorner)$ (*) for the formula ψ. But if $T(\ulcorner\psi\urcorner)$, then ψ by the unrestricted T-schema, so by (*), we have $\neg T(\ulcorner\psi\urcorner)$. And if $\neg T(\ulcorner\psi\urcorner)$, then by (*) we have ψ, so by the T-schema we have $T(\ulcorner\psi\urcorner)$. By the disjunctive syllogism, $T(\ulcorner\psi) \wedge \neg T(\ulcorner\psi\urcorner) \rightarrow \bot$. So we have $\bot$.

The sentence $\psi : \psi \leftrightarrow \neg T(\ulcorner\psi\urcorner)$ is colloquially referred to as the *liar sentence* or simply the *liar*, meaning "this sentence is false". The liar is *paradoxical* in the sense that from classical inferences, the sentence can be shown to be both true and false.[3] Paradoxes arising from the existence of a truth predicate T with respect to which the T-schema is unrestricted are said to be *semantic paradoxes*.

One response to the proof of Tarski's theorem has been to abandon having a single truth predicate at all to deal with questions of natural language or philosophy. But there have been numerous ways to at least undermine a step of this argument, and attempt to develop another truth theory with a single truth predicate:

1. Deny certain instances of "$T(\ulcorner\phi\urcorner) \leftrightarrow \phi$" however it has been formulated.
2. Claim that the truth predicate T is *partial* and is only defined for certain ϕ.
3. Claim that the formula ψ is neither true or false, and thus, is a counterexample to excluded middle.
4. Claim that from $a \vee b$, we do not necessarily have either a or b.
5. Deny the disjunctive syllogism (and thus, modus ponens for material implication) in certain instances.
6. Deny the transitivity of deduction in certain instances.

The first two ways involve a restriction of the T-schema, at least insofar as it works in material implication. The third, fourth, and fifth ways involve using a truth theory that is not classical. The sixth way involves a change in logical structure. Given a type-free truth predicate, it may be tempting to simply claim that the T-biconditionals are not derivable unrestrictedly, without sacrificing structure and classical rules for the truth theory (so not making moves 3-6). But as we shall see, the first way is not sufficient on its

[3] According to Quine in [19] this is an antinomy.

own:

Theorem (McGee): Let L be a countable first-order language which includes the language of PA (Peano Arithmetic) and a predicate T acting on names of sentences and a predicate N such that Nx if and only if $x \in \mathbb{N}$. Let the truth theory be a set of sentences of L which:
(a) contains axioms of the theory of PA with quantifiers bounded to $\mathbb{N}$;
(b) is closed under first-order consequence;
(c) contains $T(\ulcorner\phi\urcorner)$ whenever it contains ϕ;
(d) contains all instances of the following schemata:
(1) $T(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow (T(\ulcorner\phi\urcorner) \rightarrow T(\ulcorner\psi\urcorner))$;
(2) $T(\ulcorner\neg\phi\urcorner) \rightarrow \neg T(\ulcorner\phi\urcorner)$;
(3) $\forall x : (Nx \rightarrow T(\ulcorner\phi(\dot{\bar{x}})\urcorner) \rightarrow T(\ulcorner\forall x : (Nx \rightarrow \phi(x))\urcorner)$
Then the truth theory is ω-inconsistent.

Proof: By Gödel's Diagonal Lemma, there is a predicate F of the language of PA such that $\forall y : \forall z : F(0, y, z) \leftrightarrow y = z$ and $\forall x : \forall y : \forall z : (Nx \rightarrow (F(s(x), y, z) \leftrightarrow y = \ulcorner\forall y : (F(\dot{\bar{x}}, y, \dot{\bar{z}}) \rightarrow T(y))\urcorner))$ are in the truth theory along with a sentence σ such that $\sigma \leftrightarrow \neg\forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\sigma\urcorner) \rightarrow T(y)))$ is in the truth theory.[4] From (b) we have $\neg\sigma \rightarrow \forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\sigma\urcorner) \rightarrow T(y)))$ and thus $\neg\sigma \rightarrow \forall y : (F(0, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$. But from the definition of F and (b) we have $\forall y : F(0, y, \ulcorner\sigma\urcorner) \leftrightarrow y = \ulcorner\sigma\urcorner$, so bringing our results together we have $\neg\sigma \rightarrow T(\ulcorner\sigma\urcorner)$.

Now from the definition of σ and (b) we have $\sigma \rightarrow \neg\forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$ in the truth theory and from (c), $T(\ulcorner\sigma \rightarrow \neg\forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\bar{\sigma}\urcorner) \rightarrow T(y))\urcorner)$. From (d) we have $T(\ulcorner\sigma\urcorner) \rightarrow T(\ulcorner\neg\forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\bar{\sigma}\urcorner) \rightarrow T(y))\urcorner)$, and recalling that $\neg\sigma \rightarrow T(\ulcorner\sigma\urcorner)$, we then have $\neg\sigma \rightarrow T(\ulcorner\neg\forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\bar{\sigma}\urcorner) \rightarrow T(y))\urcorner)$, so by (d) $\neg\sigma \rightarrow \neg\forall x : (Nx \rightarrow T(\ulcorner\forall y : (F(\dot{\bar{x}}, y, \ulcorner\bar{\sigma}\urcorner) \rightarrow T(y))\urcorner))$.

But then from the definition of F and (b) we have $\forall x : \forall y : (Nx \rightarrow (F(s(x), y, \ulcorner\sigma\urcorner) \leftrightarrow y = \ulcorner\forall y : (F(\dot{\bar{x}}, y, \ulcorner\bar{\sigma}\urcorner) \rightarrow T(y))\urcorner))$, so $\neg\sigma \rightarrow \neg\forall x : (Nx \rightarrow (F(s(x), y, \ulcorner\sigma\urcorner) \rightarrow T(y)))$ and indeed $\neg\sigma \rightarrow \neg\forall x : (Nx \rightarrow (F(x, y, \ulcorner\sigma\urcorner) \rightarrow T(y)))$. But from the definition of σ, we have $\neg\sigma \rightarrow \sigma$. In any case, we conclude that σ, so $T(\ulcorner\sigma\urcorner)$, so by the definition of F again we can infer $\forall y : (F(0, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$.

[4]The sentence σ means "not all results of prefixing T's to the name of this sentence are true".

The above was the base case of an inductive argument for $\forall y:$ $(F(n, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$ for each n. Now suppose that we have $\forall y:$ $(F(k, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$. Then $T(\ulcorner\forall y : (F(\bar{k}, y, \ulcorner\bar{\sigma}\urcorner) \rightarrow T(y))\urcorner)$, so $\forall y :$ $(F(k+1, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$. We have thus shown by mathematical induction that for any $x \in \mathbb{N}$, $\forall y : (F(x, y, \ulcorner\sigma\urcorner) \rightarrow T(y))$. However, we also have $\neg\forall x : (Nx \rightarrow \forall y : (F(x, y, \ulcorner\sigma\urcorner) \rightarrow Ty))$ so the truth theory is ω-inconsistent.

In particular, in a given deductive theory of truth: if every theorem of the truth theory is true, the outer and the inner logic coincide and are classical, and truth is compositional and represented by a single untyped predicate, then the truth theory has no standard models of arithmetic. After all, assuming (b), compositionality entails (1) and (2) of (d), every theorem of the truth theory being true is just (c), and the two together entail (3) of (d). That a theory of truth should have standard interpretations of arithmetic as models is surely essential, so McGee's result can be seen as a restriction on classical theories of truth just like Tarski's.

There has been a temptation in the deflationary truth literature (see for example the work of Paul Horwich in [50]) to say that any maximal consistent subset of sentences of the T-schema will do in pursuing the aims of A4. But another result by McGee in [51] shows that there are so many possible, mutually incompatible maximal consistent sets of sentences that there is nothing to be gained from such an analysis.

2.3 Why do we want a truth theory?

In the last section, it was seen that there are limitations to what can be proven in a non-trivial truth theory: either certain classical rules must be sacrificed, or the T-schema biconditionals must be restricted along with features such as compositionality. In light of these limitations, there have been some arguments that truth theories can or should be avoided as part of some of the programs A1-A4 (in particular A1, A2, and A4. We briefly address two of these arguments, outlined below:

1. Various: the "ordinary notion" of truth is incoherent, so a descriptive account of natural language semantics should avoid it, namely A2. (Some have proposed alternatives to a truth predicate to put in its place, such as Kevin Scharp in [16].)

2. Dorothy Grover (from [17]): The liar sentence is insignificant as a threat to A4 (and presumably A1 and A2), much like division by zero is insignificant as a threat to arithmetic; the liar sentence is illegitimate or not meaningful, and "liar cases" should be ruled out much like division by zero.

First, a look at the first argument. We adopt an agnostic stance on whether the *ordinary notion* of truth, i.e. that which is used in the wild by people who are not logicians, is coherent. Certainly if people accept enough classical logic axioms and structural rules at the same time as an unrestricted T-schema, then everything is true and false.[5] But a coherent truth predicate may still be useful as a *device of generalisation* in the sense of Quine. What is meant by this is that, given an infinite collection of statements, we want a means of saying that all of them are the case. This is a necessary part of A2, even if it doesn't correspond to a (trivial) ordinary notion of truth.

As for the second argument, a logical truth theory is necessary for deflationists to make sure that their notion of truth is coherent or sensible whatever the empirical circumstances. An analogy is made between excluding division by zero in integral domains. The analogy might well hold to the extent that paradoxical sentences are not meaningful, though we do not have to take a particular stance on this matter - there is still a crucial disanalogy. Whenever a division by zero is made in a failed mathematical proof, it is not to be blamed on unfortunate circumstances but rather purely on human error. Division by zero can be identified by a careful reading of the proof in isolation. However, *contingently paradoxical* sentences, discussed in the next section, may be perfectly legitimate in certain circumstances and not identified as paradoxical in isolation. So there is no quick and easy way to get around the semantic paradoxes, and a good truth theory is necessary to at least avoid disaster if not diagnose them perfectly.

The next section will cover Kripke's theory of truth, an influential work which is comparable in one way or another to each of the four theories to be investigated in this thesis. It will evade the difficulties raised by the theorems of Tarski and McGee, and also the perils of typed truth predicates (it will be a type-free truth theory).

[5] This has been argued by Hartry Field.

2.4 Kripke's theory of truth

In this section we cover Kripke's theory of truth, a truth theory that is type-free and thus evades the *Nixon/Dean problems* faced by typed theories that we will mention shortly. It forms the basis of most of the theories covered in the later parts of this thesis, and is at least comparable to all of them, so it is worth at least fixing notation. It may first be prudent to first provide a note on why it has become so central.

Recall that Tarski's own approach to addressing his own negative result was to replace the problematic type-free notion of truth with a typed notion. But the typed notion of truth was decisively attacked by Saul Kripke in [47] with so-called Nixon-Dean problems showing them to be unable to cope with peculiar empirical circumstances.

The particular example that Kripke gives is along the lines of a situation - what is said about the Watergate affair - spoken about by two participants, Nixon and Dean. What Nixon and Dean say about the Watergate affair itself belongs to the set of things said about the Watergate affair. This kind of situation is encountered fairly often in everyday speech. Yet a seemingly unproblematic sentence, when uttered by Nixon, like:

"Most (i.e. a majority) of the things said by Dean about Watergate are true."

is paradoxical when Dean says:

"Everything Nixon says about Watergate is false."

and exactly half of the other things Dean says about Watergate are true (and the other half false), and every other thing Nixon says about Watergate is false. When truth is considered to be typed, then Dean must "choose" a subscript higher than Nixon in order to say what he intends to say, which is not a reflection of what happens or should happen in reality.

We now move on to Kripke's type-free proposal. We shall, from a T-free ground model $M*$, construct a model M of the language L. The language L includes a truth predicate T with an extension S_1 and an anti-extension S_2. T and hence L is defined by transfinite induction from models M_α with corresponding languages L_α. M_0 corresponds to the ground model, but with

a predicate T with a certain stipulated extension and anti-extension. Given M_α, the model $M_{\alpha+1}$ is such that T has its extension $S_1^{\alpha+1}$ as the set of names of true sentences of L_α, and its anti-extension $S_2^{\alpha+1}$ being the set of names of false sentences of L_α and of elements of D that are not names of sentences of L_α. Finally, for a limit ordinal θ, L_θ is such that $(S_1^\theta, S_2^\theta) = \bigcup_{\alpha<\theta}(S_1^\alpha, S_2^\alpha)$.

Now note that given an appropriate choice of model M_0 and valuation scheme for M, the function ϕ sending (S_1^α, S_2^α) to $(S_1^{\alpha+1}, S_2^{\alpha+1})$ for any α is *monotonically increasing* with respect to an order $\leq'$ such that $(a,b) \leq' (c,d) \leftrightarrow a \subseteq c \wedge b \subseteq d)$. Given that ϕ is monotonically increasing, there is a fixed point theorem:

Theorem: If ϕ is monotonically increasing, there exists a limit ordinal θ such that $(S_1^\gamma, S_2^\gamma) = (S_1^{\gamma+1}, S_2^{\gamma+1})$ for all $\gamma \geq \theta$.

For the valuation of M, the Strong Kleene valuation scheme is typically (but not necessarily) used. In this scheme, for some proposition ϕ one has $\neg\phi$ true if ϕ is false, $\neg\phi$ false if ϕ is true, and undefined in case ϕ is undefined; for some other proposition ψ one has $\phi \wedge \psi$ true if both ϕ and ψ are true, false if either ϕ or ψ is false, and undefined otherwise; one has $\phi \vee \psi$ true if either ϕ or ψ is true, false if both ϕ and ψ are false, and undefined otherwise; one has $\exists x : \psi(x)$ true if ψ is true at x, false if ψ is false for all x, and undefined otherwise; one has $\forall x : \psi(x)$ true if ψ is true for all x, false if ψ is false for some x, and undefined otherwise.

Now M can be defined simply as the fixed point model M_θ in this hierarchy. The question remains as to which sentences should be declared true. One option Kripke favours for doing this is to consider the true sentences to be exactly those that are declared true at the *minimal fixed point* in the Strong Kleene valuation scheme. This is the fixed point that always exists in the Strong Kleene valuation scheme when the truth predicate in M_0 is stipulated to have empty extension and anti-extension. However, one can obtain variations of this result with different valuation schemes and different fixed points.

In the minimal fixed point the true (false) sentences are said to be *grounded true (false)*. A fixed point is said to be *intrinsic* if all of the truth values it assigns to sentences are the same as the truth values that all the other fixed points assign to the same sentences. A sentence is said to be *intrinsically true (false)*, if it is true (false) at an intrinsic fixed point.

Sentences which are neither true nor false at any fixed points are said to be *paradoxical*.

One of the greatest strengths of Kripke's theory of truth with respect to the Strong Kleene valuation scheme is that it is *conservative* with respect to models of arithmetic: the predicate T can be introduced to any theory of arithmetic without affecting the truth or falsity with respect to a given model of any T-free formula in the theory. So there is no lingering fear of inconsistency. For this reason, all theories covered later in this thesis use similar means of protecting themselves from paradox.

2.5 Threats to the T-schema that this thesis doesn't cover

In this thesis truth predicates are construed as acting on coded *names* of sentences in what is usually a first-order language. The principal threat to the T-schema for first-order sentences is that covered by the formal limitations of truth theories such as the theorems of Tarski and McGee. However, depending on the application of the truth theory, the first-order sentences may be sentences translated from natural language, which leads to a few other issues raised in print:

B1 Indexicals of time, subject, object, or setting: "it is raining", "we are Europeans", "you are European"...

B2 Vague predicates: "France is hexagonal"...

B3 Syntactic issues: Jaakko Hintikka's example "if any corporal can become a general, then 'any corporal can become a general' is true"; see [45].

These three categories of "counterexamples" to the T-schema are all difficulties faced in translating natural languages into formal languages. The first two difficulties, associated with *contextualism* and discussed in [46], most obviously arise if the truth predicate acts on sentences ϕ rather than the propositions they express. If Bob, an American speaking on behalf of a group of Americans and about a group of Europeans who remark of themselves that "we are Europeans", then " 'we are Europeans' is true" is true when said by Bob in regards to the statement by the Europeans. But if Bob were to say "we are Europeans" he would be uttering a falsehood. A

cartographer may accept a remark in a discussion among uneducated men that France is hexagonal is true, but in another context deny that France is hexagonal. Much has been written on the significance of the first two counterexamples, and it would be too much of a detour from our main point of interest to discuss them further. However, causes of optimism in addressing B1 and B2 have been raised by the work of Andjelkovic and Williamson in [44], and their insights can be applied to truth theories later on.

The third example is of a somewhat different nature, dealing with the effect of shifting "any" in a sentence in English. "If any corporal can become a general..." is an antecedent of an existential form, and is obviously true: some corporal has become a general, so (in the absence of radical changes to the military) some corporal can become a general. But "any corporal can become a general" is of universal form, and is almost certainly false; any corporal about to retire from service is a reasonable counterexample to draw. So we have a counterexample to one very informal version of the T-schema. The lesson to draw from B3, from James Klagge in [15], is that implication $a \rightarrow b$ is at best an approximation of "if a, then b" sentences. The sentence "any corporal can become a general implies 'any corporal can become a general' is true" is unproblematically true, and has a different logical form from "if any corporal can become a general, then 'any corporal can become a general' is true". Enough care to distinguish the T-schema from a clumsier "if a then a is true, and if a is true then a" allows one to evade Hintikka's purported counterexample.

More pressing are the negative results of Tarski and McGee: the former casting into doubt the viability of having a truth predicate with the T-schema satisfied, and the latter casting into doubt the viability of having a truth predicate at all. These are difficulties immediately raised in the formal language itself, rather than from attempting to translate sentences in natural language to formulas. The impact of these difficulties is a central theme of this thesis.

In the next section, we categorise the theories of truth we are interested in by their accordance - or otherwise - with a range of criteria proposed by Hannes Leitgeb in [14].

3 Criteria for truth theories

3.1 Leitgeb's criteria

Here we look at a range of criteria from [14]. It is impossible for any theory of truth to satisfy all of them, on pain of triviality, due to the theorems of Tarski and McGee. But a basis for comparison may appropriately begin from seeing which criteria are sacrificed.

The four theories under investigation will all have the following features:

C1 They include truth in the form of a predicate.

C2 The predicate denoting truth will have no type restrictions.

C3 With respect to this untyped predicate, every sentence in the (empirical or mathematical) theory it is applied to is true.

C4 The theory of truth should allow for standard interpretations of arithmetic.

What makes these criteria important? C1 was our goal all along; C2 was shown to be necessary through Kripke's Nixon-Dean examples; C3 and C4 are essential if we mean our truth theory to be useful. These are four of the eight criteria for theories of truth to ideally satisfy in Hannes Leitgeb's [14]. The other four criteria are as follows:

C5 T-biconditionals should be derivable unrestrictedly.

C6 Truth should be compositional.

C7 The logic of the truth theory and of its inner theory are the same.

C8 The logic of the truth theory is classical.

By Tarski's undefinability theorem, it is impossible for any theory of truth to have features C1-C3, C5, and C8. By Vann McGee's theorem in [18], it is impossible for any theory of truth to have features C1-C4 and C6-C8. Assuming C1-C4 then, we have three maximal possibilities left for a consistent truth theory:

(a) C1-C4, C6, C8

(b) C1-C4, C7-C8

(c) C1-C7

Each of the above have been satisfied by certain truth theories already in the literature.

The above is to say that we are left with, in effect, three choices: to sacrifice the T-biconditionals and the transparency of truth, to sacrifice the T-biconditionals and the compositionality of truth, or to have a non-classical theory of truth. Each of these sacrifices have been made with some justification in print by one philosopher or another, so it would be question-begging to criticise them on the basis of sacrifices made. In the next section, we will propose an alternative list of three criteria by which to judge the theories under investigation, which may cast some light on how each should be treated in future study.

3.2 Some more criteria

In discussing the four proposals, we shall be evaluating their merits against three criteria

D1 Having to be a *theory of paradox diagnosis* - Is it necessary for the proposal to provide a means of identifying the paradoxical sentences? If so, is it able to distinguish paradoxical sentences from unproblematic ones?

D2 Avoiding *revenge problems* - Can we introduce a new predicate into the language from which we can build a "revenge" of the liar paradox, and from which we have explosion?

D3 Avoiding charges of *ad hoc* approaches - Is there a motivation for the proposed change in logic other than simply to solve the problem?

Here we shall discuss the importance of each of the three criteria being chosen.

Regarding D1, considering the threat of truth-paradoxicality addressed merely by a solution to the liar paradox is often a mistake. Kripke's noted *Nixon/Dean examples* from [47] show that some contingently paradoxical sentences can be not intrinsically problematic, unlike the liar, but fall to paradox under unfortunate circumstances. So there is no intrinsic criterion

for dealing with the threat of truth-paradoxicality.

To make explicit the difficulties faced by an inappropriate treatment of D1, consider a logic with the language and axioms of ZF set theory and a truth predicate T, which is intended to provide a logical form for sentences in natural language. The liar, along with other sentences deemed "paradoxical" due to some inherent property such as being self-referential, is assigned truth value u.

Now let T be a set of five sentences. Let σ be the sentence that says "at least three of the sentences in T are true". As it happens, two of the sentences in T are truths, two are falsehoods, and one is the sentence ϕ which says "σ is false". If σ is true, then ϕ is false, but then the majority of the sentences in T are false, so σ is false. If σ is false, then ϕ is true, but then the majority of the sentences in T are true, so σ is true. Since σ has not already been accounted for (in that it would, under many choices of T, be unproblematic), the circumstances create a contradiction and, with enough of the rules of classical logic intact, they cause explosion as well.

Kripke's theory of truth is an important precursor to the theories of Field and Priest. This approach still has some limitations. The means for treating certain sentences as truth-paradoxical and others as unproblematic is nontrivial. In particular, Kripke's theory itself falls into the "gap arguments" of Anil Gupta in [36], with respect to a variety of choices of fixed point and valuation scheme: it produces sentences which have truth value u which ought to be true or false, relative to a certain valuation scheme and fixed point. This in itself is not such a big deal on its own; one might ask, why not just choose a suitable fixed point? A sentence, after all, is only Kripke-paradoxical if it is paradoxical relative to all fixed points. But under the Strong Kleene valuation scheme, it falls into another gap argument in [35] with respect to any choice of fixed point: the sentence $l \rightarrow l$, where l is the liar sentence, is not true. After all, l is Kripke-paradoxical, and any truth-functional combination of Kripke-paradoxical sentences is paradoxical. The more sensible choices of supervaluation scheme that Kripke proposes do indeed save the classical tautologies and contradictions, but they do so at the cost of compositionality.

The ambition here is that *no more sentences may be classified as paradoxical than are necessary.* We may come to see this as an unattainable goal, but it is worth setting truth theories against this standard. If a truth

theory fails to account for a certain true (or false) sentence as being true (false), then certainly it hasn't provided the meaning of the sentence as in A1, and hasn't provided an exhaustive meaning of what it is for a sentence to be true, and thus inadequate for A2 or A4. So the ambition of D1 is two-way: to diagnose the paradoxes (on threat of triviality), and also to not provide additional false diagnoses (otherwise it's insufficient).

Now for justifying D2 as a criterion. A note should be made first on what we mean by "revenge". There are many solutions to the liar paradox that involve introducing a new semantic notion (such as "groundedness" or "stability") or a non-classical inner theory (one where certain sentences are neither true nor false, or both true and false), in order to avoid triviality. Often there are metatheoretic notions such as "neither true nor false" that are not introduced as predicates in the object language - and if they were, then a *revenge liar* would exist and lead to explosion.

For example, take Kleene's strong three-valued logic, equipped with the T predicate with the unrestricted T-schema, applied to the paradoxes; ignoring the issues raised by D1 for now, treat the paradoxical sentences as having truth value u. Introduce a new predicate N interpreted as "nontruth"; for any sentence ϕ, $N(\ulcorner\phi\urcorner)$ is true if and only if ϕ is not interpreted to have truth value 1. Let ψ be the sentence $N(\ulcorner\psi\urcorner)$, or "this sentence is nontrue". If ψ is true, then $N(\ulcorner\psi\urcorner)$ is false, so ψ is false, but then $N(\ulcorner\psi\urcorner)$ is true. If ψ has truth value u, then $N(\ulcorner\psi\urcorner)$ is true, so ψ is true, so $N(\ulcorner\psi\urcorner)$ is false. If ψ is false, then $N(\ulcorner\psi\urcorner)$ is true, so ψ is true, so $N(\ulcorner\psi\urcorner)$ is false. In any case $N(\ulcorner\psi\urcorner) \wedge \neg N(\ulcorner\psi\urcorner)$, from which we have explosion. The sentence $\psi : N(\ulcorner\psi\urcorner)$ is the *revenge liar* in this context.

In the above example, "neither true nor false" is clearly a coherent notion. An inability to express it in the truth theory suggests that it lacks some explanatory power. This would be an obstacle for pursuing the ends of Russell's philosophical logic program A2, at the very least.

Finally, there is the issue of D3, which leads to a tangle of philosophical argumentation. Many obscure changes in logic have been made with the ambition of ensuring that the introduction of an unrestricted T-schema does not lead to explosion. In order to decide whether any of these theories are acceptable, we are left with a decision as to whether the unrestricted T-schema is more desirable to keep than some rule of classical logic, or some property of the connectives of classical logic. For example, if Kripke's theory

of truth with the Strong Kleene valuation scheme is implemented, paradoxical sentences cannot imply themselves. In this case, the theorist should have a good argument for implementing their change in logic, and for the consequences (e.g. for the reflexivity of the conditional, as with applying the Strong Kleene valuation scheme to Kripke's theory). Unfortunately, this is often not the case.

Frequent changes to increasingly complicated theories of truth for the sake of avoiding problems pose great difficulties. Ad hoc methods are unscientific and have no place in a serious theory of semantics, so would take away from A1 and A2. A good defense of the deflationary position by means of a nonclassical truth theory as in A4 should also come with a good justification for violating certain classical rules.

There are a few lessons to draw here. First, the proposed logic should come with a means of diagnosing all of the truth-paradoxical sentences. Second, truth-paradoxical sentences cannot merely be treated as having a separate truth value from true and false; it is not necessarily the wrong way to go, but the threat of revenge makes the construction of such a theory a nontrivial matter. And third, every change in the logical rules and the properties of the connectives should have some justification apart from being able to cope with the liar paradox.

Having shown that D1-D3 are prescient demands, in the later sections we shall see how the four accounts of truth and paradox cope with them.

4 Paracomplete truth theories - Field's theory of truth

4.1 Leading in

A *paracomplete* truth theory is one where excluded middle is not generally valid in the logic of the truth theory for sentences including the truth predicate T. This is the first of two kinds of non-classical truth theory that will be considered here, the other being the paraconsistent truth theories. Many paracomplete truth theories have emerged over the years, though the one being investigated will be Hartry Field's from [41]. Other examples include Kripke's theory of truth with a Strong Kleene valuation scheme, and the continuum-valued logic of Jan Lukasiewicz developed in [7] applied to truth theories. There are three reasons why Field's logic takes priority here, and in this opening section I will provide them.

First, Field's truth theory satisfies a maximal consistent sublist of C1-C8, namely C1-C7. As unlike Kripke-Strong Kleene, which is perhaps the most famous of the lot, the T-biconditionals are derivable unrestrictedly. And the inadequacy proof for continuum-valued semantics in [8] shows in effect that the continuum-valued Lukasiewicz truth theory is ω-inconsistent, and thus cannot satisfy C4.

Second, Field's truth theory comes with some welcome intersubstitutivity and conservativeness results. It can show more than transparency for the truth predicate T; it is even the case that for any sentences C and D, if C and D are alike except where C has "A", D has "$T(\ulcorner A \urcorner)$", one can legitimately infer C from D and D from C.

Third, Field's truth theory is a culmination of some of the more important paracomplete truth theories that came before - in particular, aspects of Kripke's theory of truth and Lukasiewicz's theory of truth are both applied in Field's theory, and their respective difficulties (mostly) overcome.

Next, this theory shall be summarised, and then evaluated against the three criterion D1-D3. Field claims that his solution is revenge-immune, but I claim that it is not. Most of the other criticisms revolve around the complicated nature of Field's new form of implication.

4.2 Field's theory of truth - background

The principle ideas of Field's theory of truth, gradually built up over accounts in [39], [40], and [43], eventually culminating in [41], draw from Kripke's theory of truth. A notable feature of Field's theory of truth is the *Intersubstitutivity Principle* (IP): if C and D are alike except where C has "A", D has "$T(\ulcorner A \urcorner)$", one can legitimately infer C from D and D from C. Another notable feature is the conservativity of Field's theory of truth. Another is the new inclusion of an operator indicating that a sentence is *determinately* the case.

A strong enough form of implication may in the presence of intersubstitutivity of truth and the T-schema lead to *Curry's paradox*, a consequence of equipping transparency of truth with "too many" classical rules for implication: namely the deduction theorem and modus ponens. Curry's paradox can arise in a few different ways, so we shall make it explicit. We know that we can construct a sentence k which is the same as $T(\ulcorner k \urcorner) \rightarrow 0 \neq 0$. If we are to assume $T(\ulcorner k \urcorner)$, then by the transparency of T we can assume k which is just defined to be $T(\ulcorner k \urcorner) \rightarrow 0 \neq 0$. Then from modus ponens we can infer $0 \neq 0$. So we have proved $T(\ulcorner k \urcorner) \vDash 0 \neq 0$. By the deduction theorem, then, $\vDash T(\ulcorner k \urcorner) \rightarrow 0 \neq 0$, so $\vDash k$ and thus by transparency $\vDash T(\ulcorner k \urcorner)$. But then by our earlier reasoning $\vDash 0 \neq 0$ - or really, anything we could ask for.

Here the *ground language* L is a T-free language that is rich enough to express its own syntax, and the *extended language* L^+ is the result of adding T to L. Negation, conjunction, disjunction, and the quantifiers have the Strong Kleene semantics assigned to them, and none of these are defined from our new form of implication. Kripke's minimal fixed point construction is then applied to obtain a classification of the true, false, and u-valued (or in this case, 1/2-valued) sentences - apart from those containing the new form of implication.

The new form of implication is instead defined in such a way that it has a *transparent valuation*, that is that for any statement C written as one statement implying another, if D is the result of replacing some subsentence A in C with $T(\ulcorner A \urcorner)$ then D and C have the same truth value. This assures that the whole set of sentences constituting the extension of the minimal fixed point obeys transparency. An appropriate choice of transparent valuations for sentences with implication is a nontrivial matter, and indeed this is managed by a determination of each truth value akin to that of the revision

theory of truth.

The revision process for implication is as follows:

Given d an assignment function,
$|\phi \rightarrow_f \psi|_{d,0} = 1/2$;
$|\phi \rightarrow_f \psi|_{d,\alpha+1} = 1$ if $|\phi|_{d,\alpha} \leq |\psi|_{d,\alpha}$;
$|\phi \rightarrow_f \psi|_{d,\alpha+1} = 0$ if $|\phi|_{d,\alpha} > |\psi|_{d,\alpha}$;
and if λ is a limit ordinal...
$|\phi \rightarrow_f \psi|_{d,\lambda} = 1$ if $(\exists\beta < \lambda)(\forall\gamma)(\beta \leq \gamma < \lambda \rightarrow_f |\phi \rightarrow_f \psi|_{d,\gamma} = 1)$;
$|\phi \rightarrow_f \psi|_{d,\lambda} = 0$ if $(\exists\beta < \lambda)(\forall\gamma)(\beta \leq \gamma < \lambda \rightarrow_f |\phi \rightarrow_f \psi|_{d,\gamma} = 0)$;
$|\phi \rightarrow_f \psi|_{d,\lambda} = 1/2$ otherwise.

Now say that for a truth value n, $< \phi, d >$ has *ultimate value* n whenever there is an α such that for all $\gamma \geq \alpha$, $|\phi|_{d,\gamma} = n$. The semantics of the logic is given by taking $|||\phi|||_d$ to be 1 whenever $< \phi, d >$ has ultimate value 1, taking $|||\phi|||_d$ to be 0 whenever $< \phi, d >$ has ultimate value 0, and taking $|||\phi|||_d$ to be $1/2$ otherwise. Every sentence has an ultimate truth value, even if as we shall see paradoxical sentences are seen as being not *determinately* true. By the *Fundamental Theorem* of Field's theory, for any ordinal μ there are ordinals $\nu > \mu$ such that for every formula ϕ and assignment function d, $|\phi|_{d,\nu} = |||\phi|||_d$. Such an ordinal ν is called an *acceptable ordinal*.

In order to ensure transparency in the face of strengthened liars (e.g. "this sentence is not determinately true") and strengthened Curry sentences (e.g. "if this sentence is true, then 'if this sentence is true, the earth is flat'"), new truth values can be introduced into the language; and, indeed, infinitely many. This can be achieved by a generalised semantics that assigns a value to each sentence according to the set of values that it receives from the revision process.

First we have the following constraints on the value space $< V, \leq_V >$:
$\leq_V$ is a partial order of elements of V; V forms a *deMorgan lattice* with respect to operators corresponding to conjunction, negation, and disjunction
the value 1 is *join-irreducible*;
the negation operator leaves the semantic value $1/2$ fixed;
the value of $\forall x : A(x)$ is the greatest lower bound of the values of each $A(u)$;
the value of $\exists x : A(x)$ is the least upper bound of the values of each $A(u)$;

Let Δ_0 be the smallest acceptable ordinal, and let Σ be the smallest

initial order greater than Σ_0, itself the smallest acceptable ordinal that is strictly greater than Δ_0 (Σ will be a right-multiple of Σ_0 and also be acceptable). Now define V to be the set of functions f from the set of ordinals smaller than Σ to $\{0, 1/2, 1\}$ such that: if $f(0) = 1$ then $\forall\alpha : (f(\alpha) = 1)$; $f(0) = 0$ then $\forall\alpha : (f(\alpha) = 0)$; $f(0) = 1/2$ then there exists a $\rho < \Sigma$, for which there is a δ such that $\rho \cdot \delta = \Sigma$, such that $\forall\alpha : \forall\beta : (\rho \cdot \alpha + \beta < \Sigma \rightarrow_f f(\rho \cdot \alpha + \beta) = f(\beta))$.

Furthermore, define $\leq_V$ to be the set of pairs $< f, g >$ with $f, g \in V$ where $< f, g > \in \leq_V$ if and only if $\forall\alpha < \Sigma : f(\alpha) \leq g(\alpha)$. The truth value negation operator $*$ is defined by $f*(\alpha) = 1 - f(\alpha)$, and the value of a conjunction is the minimum of the values of each of the conjuncts. The truth value implication operator $\Longrightarrow$ is defined for $\alpha \neq 0$, $\alpha < \Sigma$ by: $(f \Longrightarrow g)(\alpha) = 1$ if $(\exists\beta < \alpha)(\forall\gamma)(\beta \leq \gamma < \alpha \rightarrow_f f(\gamma) \leq g(\gamma))$; 0 if $(\exists\beta < \alpha)(\forall\gamma)(\beta \leq \gamma < \alpha \rightarrow_f f(\gamma) > g(\gamma))$; 1/2 otherwise. At 0 it is defined as if it were for Σ: $(f \Longrightarrow g)(0) = 1$ if $(\exists\beta < \Sigma)(\forall\gamma)(\beta \leq \gamma < \Sigma \rightarrow_f f(\gamma) \leq g(\gamma))$; 0 if $(\exists\beta < \Sigma)(\forall\gamma)(\beta \leq \gamma < \Sigma \rightarrow_f f(\gamma) > g(\gamma))$; 1/2 otherwise. This provides us with the necessary join-irreducible deMorgan lattice of a partially-ordered infinity of truth values.

Finally, a *determinacy operator* D is introduced, defined by $D\phi \equiv \phi \wedge \neg(\phi \rightarrow_f \neg\phi)$. The reason for introducing the determinacy operator is that the liar and numerous revenge liars can be given a similar diagnosis of their truth or falsity - the liar is neither true nor false, but it is not determinately true. Now given that L is the T-free object language under scrutiny, let an *L-path of length λ* be a function p assigning to each ordinal $\alpha < \lambda$ a formula of L that is true of α and nothing else. Iterations of the predicate can stretch on into infinity, with a sequence of operations designed to go on for as far as the object language will allow:

Let p be any L-path, then:
$D^0_{(p)}$ is the identity operator on L^+;
$D^{\alpha+1}_{(p)}$ is the operator that sends each sentence $\phi \in L^+$ to $DD^{\alpha+1}_{(p)}\phi$;
$D^{\lambda}_{(p)}$, for λ a limit ordinal, is the operator that sends each sentence $\phi \in L^+$ to a sentence χ which is true if and only if for each $\alpha < \lambda$, we have that $D^{\alpha}_{(p)}\phi$ is true;
D^{α} is the operator such that $D^{\alpha}\phi$ if and only if $\exists p : (p$ is an L-path of greater length than $\alpha \wedge T(\ulcorner D^{\alpha}_{(p)}\phi \urcorner))$.

4.3 Condition D1 and the rejection of gaps

As we've seen, Field defines his form of implication via a revision-theoretic approach implementing a hierarchy of valuations. The crucial convention, though, is this: that the starting point of the hierarchy is the Strong Kleene valuation scheme with respect to which the implication-free fragment of the language is defined, and the paradoxical sentences are evaluated relative to Kripke's *minimal fixed point*. By this means, the liar cannot be distinguished from the truth-teller, but room is left for the two to be distinguished, because Field's Fundamental Theorem holds at any fixed point: the truth teller is not determinately true at the minimal fixed point, whereas the liar is not determinately true at any fixed point.

Nevertheless, in this section we will still run into a problem:

E - Paradoxical sentences are neither true nor false, but this cannot be expressed in the theory - a notion of determinateness instead serves to replace it, but doesn't apply to all paradoxical sentences.

The new form of implication is notably stronger than Kripke's in the Strong Kleene valuation scheme. It can imply everything that implication can in Kripke-Strong Kleene, but with an added bonus: sentences like $l \leftrightarrow_f l$ in Field's theory, where l is the liar sentence, are in fact true. But one particular limitation still remains. Even though presumably no contradiction is derivable from the liar paradox, the sentence $\neg(l \wedge \neg l)$ is not true in Field's theory - and also, even though the liar is neither true nor false, the sentence $\neg(l \vee \neg l)$ is not true in Field's theory.

The above is not just an artefact of the Kripkean construction. After all, de Morgan laws and double negation laws ensure that $\neg(l \vee \neg l)$ and $\neg l \wedge l$ are equivalent. So what can be said about the liar, given Field's principles about truth, cannot always be said in the object language. What is said instead is that the liar is not determinately true, and presumably every sentence that would be declared paradoxical in Kripke's theory of truth under the Strong Kleene valuation scheme is given some diagnosis in the determinacy hierarchy - not determinately true, not determinately determinately true, or some other.

Analysis from Philip Welch in [12] shows that there are sentences whose indeterminacy or determinacy can be attested by any iteration of determinacy predicates D^{α}; Hans Herzberger's revision theory encounters these in

exactly the same way, and their treatment will be given in detail in section seven. Refer to these sentences as *sporadic*. They are of course appropriately diagnosed as paradoxical in Field's theory, but they are counterexamples to the view that any truth-paradoxical sentence is dominated by determinacy predicates in Field's determinacy hierarchy.

4.4 Condition D2 and superdeterminacy

The old paradoxes that Kripke's theory faced are also faced by that of Field. Most notable is the paradox introduced by the notion of "boolean negation" which essentially brings back excluded middle, creating the first revenge paradox of this section. This particular paradox is seen to be quite unavoidable, despite the repeated claims from Field that his theory "escapes revenge": it essentially involves introducing a new bivalent truth predicate, matching the theory's non-bivalent notion of truth, so the Tarski undefinability theorem's argument can be made again. Kripke's theory had true and nontrue; Priest's theory will be seen to have false only and not false only; Field's theory has ultimately true and not ultimately true. We elaborate on this difficulty in this section, and follow it up by describing the extent to which the determinacy hierarchy does provide a suitable response to Kripke's strengthened liar and might avoid "hyperdeterminateness" paradoxes of its own.

Field's theory still faces some of the same paradoxes that Kripke's theory faced. The most immediate here is that derived from a notion of "boolean negation". Introduce a predicate N such that $N(\ulcorner\phi\urcorner)$ if and only if ϕ does not have truth value 1. Let ψ be a sentence such that $N(\ulcorner\psi\urcorner) \leftrightarrow_f \psi$. Then we have seen that from either transparency or the T-schema, and the definitions of T, N, and ψ, we can derive a contradiction. Indeed we can think of N as "not ultimately true". The paradox is just as significant for Field's theory as it was for Kripke's.

Field's response is to question why one should assume excluded middle for any stipulated negation operator: "if, for instance, one doesnt assume excluded middle for not, then there is no way to derive from the stipulation that either x or $[N(\ulcorner x\urcorner)]$ is true". So presumably Field does not think it legitimate to reason by cases that either $T(N(\ulcorner\psi\urcorner))$ or $T(\ulcorner\psi\urcorner)$. This reasoning is suspicious: Field either denies that N is worth including in the object language, or he denies that excluded middle should hold in the model theory.

The latter is not a legitimate move, and for the former the burden of argument lies with Field. But if we just accept that truth-paradoxical sentences express propositions that are indeterminate in some sense or another, then indeed N is not relevant.

The way the sentence "this sentence is nontrue" is accounted for in Field's theory is as meaning the same as "this sentence is not determinately true", which is not determinately determinately true. Accepting this reading, Kripke's own revenge paradox is suitably accounted for by the determinacy hierarchy. It may be tempting to try to come up with a paradox that forms itself on the notion of a "hyperdeterminateness" predicate that dominates all iterations of the determinacy operator in the hierarchy.

While the first paradox comes up in Kripke's theory as well as Field's, the *hyperdeterminateness* paradox only concerns Field's in that it has to do with the determinacy operator. Let H be a predicate of sentences such that in all models:

(1) $\vDash H(\ulcorner\phi\urcorner) \rightarrow_f \phi$

(2) $\phi \vDash H(\ulcorner\phi\urcorner)$

(3) $\vDash H(\ulcorner\phi\urcorner) \rightarrow_f DH(\ulcorner\phi\urcorner)$

Now let ψ be equivalent to $\neg H(\ulcorner\psi\urcorner)$. By (1) and the definition of ψ, we have $\vDash H(\ulcorner\psi\urcorner) \rightarrow_f \neg H(\ulcorner\psi\urcorner)$. Now this must mean that we have $\vDash \neg DH(\ulcorner\psi\urcorner)$, so by the contrapositive of (3) we have $\vDash \neg H(\ulcorner\psi\urcorner)$, so $\vDash \psi$. But then by (2) we have $\vDash H(\ulcorner\psi\urcorner)$, so $\vDash H(\ulcorner\psi\urcorner) \wedge \neg H(\ulcorner\psi\urcorner)$.

Field's issue with this argument is that any H that satisfies this definition should be an "intelligible notion", and it is not obvious what this could be. The seemingly obvious candidate of $\vDash H(\ulcorner\psi\urcorner)$ if and only if for any α, we have $\vDash D^{\alpha}\psi$, does not suffice: due to the eventual break-down of the determinacy hierarchy, H has nothing in its extension. Moreover, it is impossible to specify where the hierarchy breaks down. Without an appealing notion to turn to, the "hyperdeterminacy" paradox is not so much a paradox as it is merely a set of postulates inconsistent with the notion of determinacy. Perhaps, Field argues, the problem is not with determinacy but with the additional postulates for H.

The revenge paradoxes can be placed in two categories: those that follow from accepting the transparency of truth, and those that don't. The boolean negation paradoxes were present in Kripke's theory all along, and are essentially reappearances of Tarski's undefinability theorem. The task of avoiding them via a paracomplete theory is as far away as it ever was, and this limitation may just be something to live with if the transparency of truth is to be taken as premiss. The prospective difficulties with additional determinateness predicates, on the other hand, would fall in the second category: they would suggest a shortcoming in the way the determinacy hierarchy is set up. Nevertheless, the only convincing revenge paradox found here is the former:

F - The predicate "ultimately true" cannot be introduced into the language on threat of paradox.

4.5 Condition D3 and the limits of implication

The profound consequences of extending Kripke's theory of truth with a new form of implication become very apparent with the sudden presence of Curry's paradox (and slight variations thereof). Even without excluded middle, if one has a truth predicate T with transparency, and a form of implication satisfying modus ponens and the deduction theorem, one is forced to accept contradictions. Field concludes that either modus ponens or the deduction theorem must be sacrificed in reasoning with paradoxical sentences. But the nature of Field's implication is also under question. Why use revision sequences, other than to produce favourable results? There are three outstanding issues with the form of implication chosen:

G1 - The choice of modus ponens/deduction theorem appears to be arbitrary.

G2 - Field's implication, as it is defined, lacks a scientific or otherwise non-ad hoc justification.

G3 - Field's implication is unwieldy and unpalatable.

We shall elaborate a bit more on each of these points in what follows.

First, a note on G1. Field's approach is to abandon the deduction theorem. Indeed the deduction theorem is not true in general for paradoxical sentences under the form of implication used. Field argues that to assert

a conditional (taking Field's new form of implication as that conditional) on the basis of conditional assertion one needs the law of excluded middle, and since the law of excluded middle is rejected, the deduction theorem no longer holds. But the alternatives, to limit the intersubstitutivity of truth or the structural rules of deduction, are ignored without much comment. Furthermore, it could be that there is a duality between Field's theory and a corresponding paraconsistent theory, perhaps rejecting modus ponens instead of the deduction theorem - the burden would then be on Field to argue in principle for paracomplete solutions instead of paraconsistent ones.[6] In any case, a sacrifice has been made for the sake of the other desired rules, and it is not clear whether the sacrifice is the right one.[7]

Now, on G2 and G3. There does not appear to be a particularly strong motivation for using the new form of implication apart from gearing it to the desired results for the truth theory. Certainly a logical cognitivist such as Robert Hanna in [5] must reject Field's form of implication as a replacement for material implication - according to his central thesis "logic is cognitively constructed by rational animals" but it would take literally forever for the truth or falsity (or particularly the lack thereof) of certain sentences with embedded implication operations to be constructed by such creatures. Its applicability outside of truth theories is also questionable, due to its high degree of complexity. In particular, it has been shown in [**?**] that Field's logic is not axiomatisable.

In fact, Field himself was not happy with the present form of implication for the reason that it is too weak to accommodate certain classical rules governing restricted quantification. As of this writing, current work of his is focused on replacing this with a stronger form of implication, but nonetheless one without a deduction theorem. Seeing how central the revision process is in much of his logic as it is presented in this section, we shall refrain from entering a digression on how this could be done. However, objections G1-G2 (at least) would still stand.

[6] Much of [41] is indeed dedicated to a criticism of Priest's Logic of Paradox, but it is on two fronts: that paraconsistent theories do no better than paracomplete theories in dealing with revenge (despite appearances), and that the logic of paradox lacks some of his favourite logical principles.

[7] As I understand from his writings, Field is prepared to accept logical pluralism, but mainly insofar as there may not be one sensible choice of semantics for implication with respect to which the resulting logic is strongest, leaving the syntax and semantics for his logic otherwise fixed; see [32].

5 Paraconsistent truth theories - The logic of paradox

A *paraconsistent* truth theory is one where certain sentences including the truth predicate T are considered to be both true and false. A range of paraconsistent truth theories have emerged over the years, most prominently from Graham Priest and J.C. Beall. The one being investigated will be Priest's *logic of paradox*. The logic of paradox has been frequently referenced and used as a point of comparison in truth literature, and being almost identical to the three-valued logic K_3 used as the basis of Kripke's theory of truth with respect ot the Strong Kleene valuation scheme, invites many useful comparisons to the Kripkean theories. This will be of help to us later on in investigating a duality between paracomplete and paraconsistent theories.

5.1 Background

The principle of *dialethism* that certain sentences are both true and false motivates the idea of F.G. Asenjo's early work in [30] and later Graham Priest's *logic of paradox* (LP) in [37]. In particular, the T-schema is taken as premise, and the liar sentence is allowed to produce a contradiction. However, the contradiction does not entail explosion due to certain rules of classical logic not being satisfied. This is due to the valuation scheme imposed on the logical connectives and the definition of the entailment relation.

The valuation scheme would be identical to that of the Strong Kleene valuation scheme, but rather with u replaced by p, the truth value that a sentence ϕ is said to take whenever ϕ is true and false (i.e. when it is *paradoxical*). The entailment relation is that for any set of sentences Σ, $\Sigma \vDash A$ if and only if there is no valuation v on all the sentences of L for which $v(A) = 0$ and for all $B \in \Sigma$, $v(B) = 1$ or p.

What happens with the liar is this: it is supposed that either l or $\neg l$. If l, by the T-schema $T(l)$, but this is equivalent to $\neg l$. Moreover if $\neg l$, this is equivalent to $T(l)$, and by the T-schema we have l. So far this is along the same lines as an argument leading to explosion. But in either case, l and $\neg l$ both have truth value p, and going by the meaning of entailment, $l \wedge \neg l \nvDash \bot$

which is a counterexample to explosion. Moreover, since $(l \vee \bot) \wedge \neg l \not\models \bot$ we have a counterexample to disjunctive syllogism as well. This is the double-edged sword of the logic of paradox: it saves logic from triviality in the face of contradictions, but at the cost of the disjunctive syllogism, and thus also modus ponens for material implication.

So the classically-valid rules of *modus ponens*, *modus tollens*, and *reductio ad absurdum* are no longer valid in the logic of paradox. Instead, they are *quasi-valid* in the sense that they are truth-preserving in the absence of paradoxical sentences. If a sentence ϕ is not paradoxical (a safe assumption, for our purposes, when ϕ is T-free), if $\neg\phi$ then ϕ has truth value 0; then if $\phi \wedge \psi$ (ψ also T-free) then we have ψ. But in the "paradoxical fragment" (which would depend, as discussed before, on extrinsic circumstances) reasoning would be crippled.

5.2 Condition D1 and the methodological maxim

Here we investigate the ability of the logic of paradox to handle condition (1): that the theory is adequate at diagnosing paradoxical situations. Of course, *prima facie* there is no need to diagnose the semantic paradoxes. Variably paradoxical situations can lead to contradictions without the threat of explosion. But as Priest admits, sacrificing disjunctive syllogism would amount to "crippling classical reasoning" whenever paradoxical sentences are involved. The compromise he draws in [37] is the *methodological maxim* that "unless we have specific grounds for believing that paradoxical sentences are occurring in our argument, we can allow ourselves to use both valid and quasi-valid inferences". But if we are to include some "quasi-valid" form of entailment, grafting classical reasoning onto the logic of paradox, it becomes necessary to diagnose paradoxical situations in order to avoid explosion.

Still, there is a means by which one can diagnose the paradoxical sentences. Given the similarity to the Strong Kleene valuation scheme, one can technically accommodate the Logic of Paradox via a dual approach to Kripke's construction, which we shall call *the dual construction*. The dual construction has a hierarchy of languages once again, but this time let the truth predicate have extension $R_0^+ = D$ and anti-extension also $R_0^- = D$ at L_0. At the same time the extension and anti-extension S_α^+ and S_α^- of the Kripke construction can also be defined as before, treating p like u. For the successor case of the dual construction, at $L_{\alpha+1}$ the truth predicate has

extension $R^+_{\alpha+1} = R^+_\alpha \setminus S^-_{\alpha+1}$ and antiextension $R^-_{\alpha+1} = R^-_\alpha \setminus S^+_{\alpha+1}$. When γ is a limit ordinal, $R^+_\gamma = \bigcap_{\beta<\gamma} R^+_\beta$ and $R^-_\gamma = \bigcap_{\beta<\gamma} R^-_\beta$. There are various fixed points of varying size, in particular the *maximal fixed point* that arises having taken $R^+_0 = R^-_0 = D$ at L_0.

The methodological maxim can be supplemented with the dual construction, though the construction is imperfect in its diagnosis of *non-paradoxical* situations. With respect to any fixed point, the sentence $l \to l$ is both true and false. Yet while neither l nor its negation can be used with modus ponens, $l \to l$ itself is unproblematic. This leads us to a general conclusion: that the dual construction produces more contradictions than are necessary.

Alternatively, Priest has attempted to introduce conditional operators in the logic of paradox for which modus ponens does hold. Here it is necessary to give a proof that the extended logic does not entail everything. This has been achieved in [38], but as noted in [41], not all sentences of the form $A \to (B \to B)$ will be true in the T-free fragment of LP - contradicting the requirement we assumed at the very beginning of this chapter that all classical inferences should hold in the T-free fragment of the logic in question.

5.3 Condition D2; revenge

Now we turn to revenge issues. One might at first imagine that since in LP contradictions do not entail explosion, forcing new contradictions by introducing new operators does not cause a problem. But under certain assumptions about the underlying metalanguage, namely that it is classical, the revenge problem becomes apparent. Take, for instance, a predicate N where $N(\ulcorner\phi\urcorner)$ is true if and only if ϕ has truth value 0 in the model (is "only false") and $N(\ulcorner\phi\urcorner)$ is false if and only if ϕ has truth value 1 (true) or p (paradoxical). This can be construed as "ϕ is only false". Then let ψ be equivalent to $N(\ulcorner\psi\urcorner)$. If ψ has truth value 1, then by the definition of ψ, ψ is false. So ψ is only true and false. If ψ has truth value 0, then ψ is only false, but ψ is then also true by definition. If ψ is paradoxical, then in particular it is true, so by the definition of ψ, ψ has truth value 0. So ψ is true and only false. In any case the interpretation function must send ψ to two different values at once, which is not what a function does.

So we are left with the problem of redefining the notion of what it means for a sentence in LP to be true, false, a dialethia, a dialethia and not a di-

alethia, and so on, in such a way that revenge problems do not appear to impose significant expressive limitations. This problem has apparently not yet been addressed in full detail, though Priest in [34] addresses it in part with a treatment of generalized truth values. That is, "only true and paradoxical" takes on a new truth value of its own, as does "only false and paradoxical" and "only false and only true". Compound formulas and entailment are revised accordingly. Priest's preferred path has been to stipulate that the logic of paradox should be formalised in a paraconsistent set theory (see [38]). We will discuss this possibility shortly.

5.4 Condition D3; methodological issues

Based on his writings, one would imagine that Priest would dismiss most of the discussion of LP here out of hand. Work of his continues on extending LP with a conditional with respect to which modus ponens holds, to avoid a commitment to either losing modus ponens or using the methodological maxim. Moreover, Priest has advocated a paraconsistent set theory in which to discuss LP; essentially a change in the way logic is done. If revenge paradoxes in an extended version of LP would lead to contradictions in set theory, then as far as Priest is concerned, that's all very well, discussing LP within his own alternative set theory.

Whether LP can be sufficiently extended with a reasonable conditional remains an issue for debate and further investigation. A few words, however, should be said about the differing treatment of revenge. We were concerned, to begin with, with accounting for a truth predicate as a part of our logical vocabulary, in common with most of the logical literature. In doing so, we agreed beforehand on the setting in which this logic was to be introduced: in classical (i.e. ZF) set theory. If we work instead in a paraconsistent set theory, we are in effect solving a different problem. The possibility of changing the way logic is done has been discussed in [31] and in more detail in [33]. We shall leave that discussion to one side in our own considerations.

At the end of it all, we are left with a mixed impression of whether Priest's LP provides a satisfactory solution to the paradoxes. The similarity of LP's valuation scheme to the Strong Kleene valuation scheme can be exploited to produce a means of diagnosing the semantic paradoxes. As we have also seen, this diagnosis of the semantic paradoxes also gives numerous false positives. The reader may note that no principled defense has been

made for the law of non-contradiction. Our concerns have been, rather, with the implicit demands to revise the fundamental laws of set theory. Instead, any paraconsistent logician should aim to derive only as many contradictions as are necessary, and should come equipped with some means of dealing with the threat of revenge paradoxes, in such a way that does not involve changing the laws of set theory. LP when equipped with the dual construction falls short in deriving only as many contradictions as are necessary, and it is doubtful that LP could deal with what are tangible revenge issues. It is perhaps more appropriately seen as a starting point for paraconsistent logics, to be extended with a new form of implication, and equipped with a new valuation scheme.

6 Theories of truth with weaker inner theories - The Leitgeb-Welch Propositional Theory

The next object of discussion will be theories of truth with *weaker inner theories* in the sense of C7. We shall primarily focus on the *Leitgeb-Welch Propositional Theory* rather than on Kripkean *classical gap* and *classical glut* theories; by [41] they do not satisfy C6.

6.1 Introduction and summary

The *Leitgeb-Welch Propositional Theory* of truth in [25], a descendent of [27] which we shall refer to as *L-W*, attempts to avoid the semantic paradoxes in the same way as ZF set theory avoids Russell's paradoxes. In L-W, we have elementhood $x \in y$ correspond to the relation y is about x. Here there is a propositional T-schema and a formula T-schema: the propositional T-schema holds for all propositions, but the T-schema for formulas only holds whenever a formula expresses a proposition. In what follows we shall formalise the notions of proposition, satisfaction, formula, and expressing a proposition, before making some remarks about how the theory stands up.

The language $L_{L\text{-}W}$ consists of the usual logical signs for a first-order logical language, along with brackets, variables, constants, and satisfaction. The predicates of the language are:

$Concat_3(x_1, x_2, x_3)$ - x_1 is the concatenation of x_2 and x_3;
$Concat_4(x_1, x_2, x_3, x_4)$ - x_1 is the concatenation of x_2, x_3 and x_4;
$Sat(x_1, x_2)$ - x_1 is satisfied by x_2;
$about(x_1, x_2)$ x_1 is about x_2;
$Tr(x_1)$ - x_1 is true;
$PropFn(x_1)$ - x_1 is a propositional function;
$Var(x_1)$ - x_1 is a variable.

The primitive individual terms and constants are NEG, CON, DIS, $IMP, EQU, UNIV$, $EXIS, ID, CONCAT_3, CONCAT_4, SAT, ABOUT$, $TR, PROPFN, VAR, X_1$, X_2... These refer to parts of propositional functions, as opposed to the variables, primitive predicates and logical signs used in the first-order language. In writing propositional functions and propositional concepts the authors use Polish notation, for example $CONCAT_3X_1X_2X_3$.

L-W has numerous axioms, which are neatly divided into two categories: those that effectively define what it means to be a concept or a propositional function (the PF axioms), and those that effectively define what it means for a propositional function to be satisfied by a choice of variable assignment (the S axioms). We only mention a few here explicitly, for most of these are close relations to (in the PF case) the ZF axioms or (in the S case) Tarski's definition of truth.

Out of the language of L-W there are *propositional functions* but also *concepts*, which make up the syntactic parts of propositional functions without being propositional functions themselves. The axioms of L-W distinguish concepts from propositional functions:

PF1: $\forall x : Concept(x) \leftrightarrow x = NEG \vee x = CON \vee x = DIS \vee ... \vee x = VAR \vee Var(x) \vee \exists u : \exists v : (Var(u) \wedge Var(v) \wedge Concat_3(x, u, v))$ (definition of concept)

PF2: $Concept(x) \rightarrow \neg \exists y : xabouty$ (concepts are not about anything)

PF11: $\forall x : Concept(x) \vee PropFn(x)$

An analogue of the foundation axiom in ZF ensures that we do not have a "liar proposition" that is the case if and only if it is not about itself:

PF5: $\forall x : (\phi[x] \rightarrow PropFn(x)) \rightarrow (\exists y : \phi[y] \rightarrow \exists y_0 : (\phi[y_0] \wedge \forall z : (\phi[z] \rightarrow \neg y_0 aboutz)))$

Propositional functions are conceived of as being syntactically built from concatenations of concepts and other propositional functions; if two propositional functions are identical, then so are their (syntactic) parts. Propositional functions can thus be uniquely specified in terms of their *conceptual form*, though they may differ in the propositions or concepts that they are about.

Analogues of the axioms of the ZFC axioms of separation, pairing, union, power set, choice, infinity, and replacement for atomic propositional functions are provided or can readily be derived. So, too, is an induction scheme for propositional functions similar to that for sets.[8] As a result, the set theory with urelements can be reconstructed from the PF axioms, defining the

[8]See an appendix at the end of this section for all of the axioms of L-W.

notion of set by $\forall x : Set_{PF}(x) \leftrightarrow Concat_3(x, TR, X_1)$ and of elementhood by $\forall x, y : x \in y \leftrightarrow Set_{PF}(y) \wedge yaboutx$.

We can build the universe V of sets inductively from the propositional function axioms. From PF5, the empty set exists. That the sets are "true" means they are grounded with respect to the aboutness relation.

The addition of set theory will be necessary in what follows. For any propositional function, refer to the set Q_z of *variable assignments* s for a proposition z as the set of relations s mapping each of the conceptual variables X that are part of the conceptual form of z to a member of the set $\{y : zabouty\}$. A valuation sequence s' is said to be a u-*alternative* of valuation sequence s with respect to a member u of s if it is identical to s but with some other variable u' in place of u. Then from the S axioms, what we have in L-W is a truth predicate on propositions that is type-free and iterable, as with Kripke's theory of truth, and also compositional.

The establishment of a truth theory without paradox requires a truth theory of not just propositions but the formulas that express them; this requires a bit more work. First, let $L^*_{L\text{-}W}$ consist of all $L_{L\text{-}W}$-formulas ϕ whose quantifiers are bounded by some Tr-free and Sat-free $L_{L\text{-}W}$-formula $\psi[x]$ - for any $L_{L\text{-}W}$-formula ϕ, let this formula in general be referred to by $\Psi(\phi)$. Let N be the set of codes of Tr-free and Sat-free $L_{L\text{-}W}$-formulas. Define Sat^- to be the satisfaction predicate of the Tr-free and Sat-free fragment of $L_{L\text{-}W}$. and for $\psi \in Fml^-$. Finally, define $b(\psi[x]) = \{n \in N \cup \mathbb{N} | Sat^-(n, \psi[x])\}$.

We also have to define the predicate $Expr(x, y)$ meaning "the formula coded by x corresponds to the conceptual form of the propositional function y". A simple rough sketch of an inductive definition shall suffice here:

for all atomic propositional functions y, $Expr(x, y)$ if and only if x is the code of $x_i = x_j$ and y is the conceptual form of IDX_iX_j for some x_i, x_j, X_i, X_j, or x is the code of $Concat_3(x_i, x_j, x_k)$ and y is the conceptual form of $CONCAT_3X_iX_jX_k$ for some x_i, x_j, x_k, X_i, X_j, X_k, or... [similarly for $Concat_4$, *about*, Sat, Tr, $PropFn$, Var, and $Expr$ itself]

for all propositional functions y of the conceptual form $NEG\Phi$, $Expr(x, y)$ if and only if x is the code of $\neg\phi$ with $Expr(\dot{\phi}, \Phi)$

for all propositional functions y of the conceptual form $CON\Phi\Psi$,
$Expr(x, y)$ if and only if x is the code of $\phi \wedge \psi$ with $Expr(\dot{\phi}, \Phi)$ and $Expr(\dot{\psi}, \Psi)$

for all propositional functions y of the conceptual form $UNIVX_i\Phi$,
$Expr(x, y)$ if and only if x is the code of $\forall x_i : \phi$ with $Expr(\dot{\phi}, \Phi)$

We then define the monotone operator ζ on sets of pairs of L^*_{L-W}-formulas and propositional formulas by: $\zeta^0 = \varnothing$; $\zeta^\lambda = \bigcup_{\alpha<\lambda} \zeta^\alpha$ for λ a limit ordinal; $\zeta^{\alpha+1} = \{< \gamma, y > | Expr(\gamma, y) \wedge \forall n \in b(\Psi(\gamma)) : ((n \in \mathbb{N} \wedge yisaboutn) \vee (n = \gamma' \wedge \exists y' : [yabouty' \wedge < \gamma', y' > \in \zeta^\alpha])) \wedge (\forall y' : [yabouty' \rightarrow y' \in \mathbb{N} \vee \exists n \in b(\Psi(\gamma))(n = \gamma' \wedge < \gamma', y' > \in \zeta^\alpha)\}$.

Now for any L^*_{L-W}-formula γ and propositional function y, $Expr_{L^*_{L-W}}(\gamma, y)$ if and only if $\zeta(\gamma) = y$. Thus we have a sort of syntactic definition of expressing. We can then say that γ is *grounded* if and only if $\gamma \in dom(\zeta)$. We can define satisfaction for L^*_{L-W}-formulas by $Sat_{L^*_{L-W}}(s*, \gamma)$ if and only if $\exists z : Expr_{L^*_{L-W}}(\gamma, z) \wedge Sat(s, z)$. Truth for formulas is then defined as $Tr_{L^*_{L-W}}(z)$ if and only if $\forall s \in S_z : Sat_{L^*_{L-W}}(s, z)$, where S_z is the set of all finite sequences s of codes of individual variables in the formula coded by z.

What is made explicit here is that only grounded formulas express propositions, and although the liar sentence can be formulated, it does not express a proposition and thus is not true in the sense given. In fact, here we see that no ungrounded sentence is true. To the theory's credit, there is a consistency result here: $Con(ZFC) \leftrightarrow Con(L - W)$.

The T-schema for formulas does not hold unrestrictedly, for not every formula expresses a proposition in L-W. We also don't have transparency for formulas: even if ϕ, we still don't necessarily have $Tr_{L^*_{L-W}}(\ulcorner\phi\urcorner)$ because ϕ may not even express a proposition. This contrasts with theories such as Field's.

6.2 Limitations of the L-W notion of groundedness

L-W's limitations reside mainly in its difficulties in considering the sentences that express a proposition as being exactly those that are grounded, from excessive limitations on the theory's notion of groundedness. We summarise these by the objections H1-H4.

The theory L-W firstly does not distinguish between paradoxes and truth-tellers: both are regarded simply as not expressing a proposition. Distinguishing between paradoxes and truth-tellers may not be the intention of Leitgeb and Welch, but it is their intention to give semantics a set-theoretic foundation, and excluding groups of sentences like those said by A and B is too restrictive for this lofty goal.

H1 - There is no means of distinguishing truth teller sentences from liar sentences in L-W.

There is a degree to which a propositional, set-theoretic analogue of Kripke's theory of truth has been obtained in L-W. In fact, the two notions match well enough that L-W falls into some of the same limitations that Kripke's (with respect to the Strong Kleene valuation scheme) would without the additional fixed points to distinguish between crucially different kinds of ungrounded sentences. Consider the following two sets of formulae:

A: $a : \forall\phi : \phi \in B \rightarrow Tr_{L^*_{L-W}}(\ulcorner\phi\urcorner)$, $b : 0 \neq 0$, $c : \exists\phi : \phi \in B \wedge Tr_{L^*_{L-W}}(\ulcorner\neg\phi\urcorner)$

B: $d : \exists\phi : \exists\psi : \exists\chi : \phi \neq \psi \wedge \phi \neq \chi \wedge \psi \neq \chi \wedge \phi \in A \wedge \psi \in A \wedge \chi \in A \wedge \neg(Tr_{L^*_{L-W}}(\ulcorner\phi\urcorner) \wedge Tr_{L^*_{L-W}}(\ulcorner\psi\urcorner)) \wedge \neg(Tr_{L^*_{L-W}}(\ulcorner\phi\urcorner) \wedge Tr_{L^*_{L-W}}(\ulcorner\chi\urcorner)) \wedge \neg(Tr_{L^*_{L-W}}(\ulcorner\psi\urcorner) \wedge Tr_{L^*_{L-W}}(\ulcorner\chi\urcorner))$, $e : 0 = 0$

This is a close analogue of Gupta's example of a dialogue between two people in [36]. Gupta's example was intended to discredit Kripke's theory of truth with respect to the minimal fixed point and the Strong Kleene valuation scheme. Thinking in terms of people and what they say, this is a collection of assertions by persons A and B: A says the statement "every sentence B says is true", some falsehood, and then the statement "there is some sentence B says that is false"; B says "there are three non-equivalent sentences said by A, and at most one of these three sentences is true", and a truth.

On the L-W analysis, b and e immediately come out as false and true respectively, being arithmetical falsehoods and truths; they are both seen to express a proposition at ζ^1. However, the proposition expressed by a is not built inductively from propositions expressed by arithmetical formulae and logical tautologies, so by the formalism of L-W there is no proposition expressed by a. Similarly with c and d. This shows a limitation of $Tr_{L^*_{L-W}}$ as a truth predicate of formulas.

There are plenty of candidates for the truth predicate, in for example Field's theory of truth and the revision theories of truth, where replacing $Tr_{L^*_{L-W}}$ by the truth predicate T under consideration, a simple argument by contradiction follows through. Let $a* - f*$ indicate the same formulas with T in place of $Tr_{L^*_{L-W}}$. Suppose that $d*$ were false. Then more than one of the sentences in A are true, so $a*$ and $c*$ are both true. But $a*$ is equivalent to the negation of $c*$, so $d*$ must be true. If we suppose that $c*$ were true, then $d*$ must be false, which we have already shown to not be the case. So $c*$ must be false, $d*$ must be true, and hence $a*$ must be true.

In Kripke's theory of truth, $c*$ is ungrounded false, and $d*$ and $a*$ are ungrounded true. The formulas are evaluated as true or false not at the minimal fixed point, but some others, and always the same statement. This applies to many groups of sentences which refer to other groups of sentences, where evaluating their truth or falsity is based on looking at consistent interpretations of the truth predicate.

H2 - Certain intuitively true sentences are considered not to express propositions in L-W.

Another difficulty concerns the truth of propositions. There might easily be circumstances where two propositional functions A and B are true, but their conjunction $CON(A, B)$ and disjunction $DIS(A, B)$ are not true: namely, where A and B are not about the same propositional functions or concepts. This arises from a restriction in PF10c of defined conjunctions and disjunctions to conjuncts and disjuncts that are about the same things; the purpose of this restriction is not clear or justified.

H3 - The condition that two propositions that are both true in L-W must nonetheless be about the same things for their conjunction or disjunction to be true is unusually strong.

The notion of groundedness in L-W is not extensionally equivalent to Kripke's under the Strong Kleene valuation scheme, however. Indeed, the formula $Tr(\ulcorner 0 = 0 \urcorner) \vee t$ where t is the liar sentence (or indeed any ungrounded sentence) is grounded in the latter, but not in the former. Thus while in the latter it is true, in the former it does not even express a proposition. Thus the semantic paradoxes cannot serve a role of deciding the truth or falsity of a conjunct; from the truth of $a \vee t$ we cannot determine the truth of a, and from the falsity of $b \wedge t$ we cannot determine the falsity of b. This, at least, deserves some justification.

H4 - The disjunction of a true sentence with a sentence that does not express a proposition - and the conjunction of a false sentence with a sentence that does not express a proposition - does not express a proposition.

6.3 A brief note on D2; revenge

In this section, we mention two cases of revenge. The first is that referred to as revenge by name in the manifesto for L-W, being an instance of a notion not expressible in the theory. However, the traditional means of obtaining revenge liars is not possible in L-W for reasons we shall mention.

Two cases of "revenge" were mentioned in presentations of L-W: the unbounded formula $\forall x : (x = x \rightarrow x = x)$ and the unrestricted liar sentence. These formulas are seen as threatening in that they are provable but do not express propositions. To argue that this is not a bad thing, Leitgeb considers the case in set theory of $x = x$ being provable but not identifiable with a set: his proposed foundation for semantics should not have to "provide its own foundation". Looking closely, we have a justification for accepting truth theories for which the outer and inner theories do not coincide: that the object of study is the inner theory. This is a coherent stance to take in pursuing the aims A1 and A2. We can see a difficulty with this stance in pursuing the aims A3 and A4, but the restriction of the T-schema has already made this impossible.

The above is not the revenge problem as we have formulated it: to derive a contradiction from the T-schema given the introduction of a new predicate. Curiously, we cannot get revenge back by the usual means. Let's say we introduce a predicate $F_{L^*_{L-W}}$ in L-W acting on any formula ϕ such that $TrL^*_{L-W}(\ulcorner F_{L^*_{L-W}}(\ulcorner \phi \urcorner) \urcorner) \leftrightarrow \neg TrL^*_{L-W}(\ulcorner \phi \urcorner) \vee \neg \exists y : Expr_{L^*_{L-W}}(\ulcorner \phi \urcorner, y)$. Then the "strengthened liar" $\psi : F_{L^*_{L-W}}(\ulcorner \psi \urcorner)$, being clearly ungrounded, can be proved by the theory. However, it is not declared true because it does not express a proposition. The advantage of having such a strong restriction on the truth of formulas is that we now have a convenient way of dismissing revenge arguments. Tarski's undefinability theorem cannot reappear because neither the necessary T-biconditionals nor transparency are there to satisfy it, so revenge in the usual sense also fails to appear.

6.4 Is the evasion of paradox in L-W ad hoc?

We briefly concern ourselves in this section with the notion of whether the evasion of paradox in L-W is ad hoc. In brief, paradox is evaded in L-W because (1) paradoxical sentences are shown to be ungrounded, and (2) ungrounded sentences do not express propositions. A convincing justification for (1) has already been shown to us; the contraposition of (1) has been shown by the relative consistency of L-W. A couple notes on (2) remain.

First, a note on whether it is contingent that sentences do not express propositions. Any contingently paradoxical sentence is diagnosed in L-W as expressing a proposition in the circumstances where it is grounded, and not expressing a proposition in the cirucmstances where it is not. Remarks have been made on the metaphysically questionable nature of having it be contingent that certain sentences express propositions. Leitgeb and Welch have responded by adopting an externalist stance on meaning, from which the consequences of their theory coherently follow.

Second, a note on whether ungrounded sentences, in general, do not express propositions. The authors maintain that they are not being ad hoc here, for they believe they are being justified by a close analogue between sets and propositions, and elementhood and aboutness. We have already expressed some doubts about whether grounded sentences in the sense of L-W are the sole truth-bearers. The particular choice of expressing relation does not come with its own justification and has some arbitrary consequences. Another objection to consider is that ZF set theory as a means of avoiding the set-theoretic paradoxes is itself ad hoc. A standard response is to note that multiple alternative foundations of set theory have been shown to be conservative extensions of ZFC.

As a foundation of semantics pursuing the aims A1 or A2, adopting the views of (a) an externalist stance on meaning and (b) the belief of a close analogy, almost identity, between the elementhood of x in y and y being about x provides us with grounds for accepting the idea behind the theory L-W. What remains mysterious are numerous artefacts and gaps of the theory, exemplified by the objections H1-H4.

6.5 Appendix: an abbreviated axiom scheme for L-W

Many of the axioms of L-W are close relations with those of ZF (the PF axioms) or Tarski's definition of truth (the S axioms), and have been skimmed over. Here we mention them for reference.

PF1: $\forall x : Concept(x) \leftrightarrow x = NEG \vee x = CON \vee x = DIS \vee ... \vee x = VAR \vee Var(x) \vee \exists u : \exists v : (Var(u) \wedge Var(v) \wedge Concat_3(x, u, v))$ (definition of concept)

PF2: $Concept(x) \rightarrow \neg \exists y : xabouty$ (concepts are not about anything)

PF3: $NEG \neq CON, NEG \neq DIS, CON \neq DIS, ...$ (no two constants denote the same concept)

PF4i: $Var(X_i)$ for all $i \in \mathbb{N}$ (X_i is a conceptual variable)

PF4a: $\neg Var(NEG), \neg Var(CON), ...\neg Var(VAR), \forall x : \forall y : \forall z : (Concat_3(x, y, z) \rightarrow \neg Var(x)), \forall x : \forall y : \forall z : \forall t : (Concat_4(x, y, z, t) \rightarrow \neg Var(x))$ (constants and concatenations are not conceptual variables)

The above axioms specify the class of *concepts*, which make up the parts of propositional functions without being propositional functions themselves.

PF5: $\forall x : (\phi[x] \rightarrow PropFn(x)) \rightarrow (\exists y : \phi[y] \rightarrow \exists y_0 : (\phi[y_0] \wedge \forall z : (\phi[z] \rightarrow \neg y_0 aboutz)))$ (foundation for propositional axioms)

PF6a: $\forall u : \forall v : \forall w : \forall x : \forall y : \forall z : Concat_3(w, u, v) \wedge Concat_3(z, x, y) \rightarrow [w = z \leftrightarrow ((u = x \wedge v = y) \wedge (PropFn(w) \leftrightarrow PropFn(z)) \wedge (PropFn(w) \wedge PropFn(z) \rightarrow \forall s : (wabouts \leftrightarrow zabouts)))]$ (identity for concatenations of two concepts)

PF6b: $\forall r : \forall t : \forall u : \forall v : \forall w : \forall x : \forall y : \forall z : Concat_4(w, u, v, r) \wedge Concat_4(z, x, y, t) \rightarrow [w = z \leftrightarrow ((u = x \wedge v = y \wedge r = t) \wedge (PropFn(w) \leftrightarrow PropFn(z) \wedge (PropFn(w) \wedge PropFn(z) \rightarrow \forall s(wabouts \leftrightarrow zabouts)))]$ (identity for concatenations of three concepts)

PF7: $\forall x : AtPropFn(x) \leftrightarrow \exists t \exists u \exists v \exists w : Var(t) \wedge Var(u) \wedge Var(v) \wedge Var(w) \wedge [Concat_4(x, ID, u, v)) \vee \exists s(Concat_4(x, CONCAT_3, u, s) \wedge Concat_3(s, v, w)) \vee \exists r : \exists s : (Concat_4(x, CONCAT_4, r, s) \wedge Concat_3(r, t, u) \wedge Concat_3(s, v, w)) \vee Concat_4(x, ABOUT, u, v) \vee Concat_4(x, SAT, u, v)) \vee Concat_3(x, TR, u) \vee Concat_3(x, PROPFN, u) \vee Concat_3(x, VAR, u)]$ (definition of atomic propositional function)

PF8-1: $\forall u \forall v : Var(u) \wedge Var(v) \rightarrow \forall x1 ... \forall xn \forall x \exists y[\forall z(yaboutz \leftrightarrow xaboutz \wedge \phi[z, x1, ..., xn]) \wedge Concat_4(y, ID, u, v)]$. (separation for atomic propositional functions with ID)

...

PF8-8: $\forall u : Var(u) \rightarrow \forall x1 ... \forall xn \forall x \exists y[\forall z(yaboutz \leftrightarrow xaboutz \wedge \phi[z, x1, ..., xn]) \wedge Concat_3(y, VAR, u)]$ (separation for atomic propositional functions with VAR)

...

PF8-41: $\forall u \forall v : Var(u) \wedge Var(v) \rightarrow \forall x_1 ... \forall x_n \forall x \exists y[\forall z(xaboutz \rightarrow \exists! t \phi[z, t, x1, ..., xn]) \rightarrow (\forall z(xaboutz \rightarrow \exists t(yabout t \wedge \phi[z, t, x1, ..., xn])) \wedge Concat_4(y, ID, u, v)]$. (replacement for atomic propositional atomic functions with ID)

...

PF8 provides analogues of the axioms of the ZFC axioms of separation, pairing, union, power set, choice, and replacement for atomic propositional functions; only a few instances here are provided as examples.

PF9: $\exists u : Var(u) \wedge \exists x[\exists y(Concat_3(y, TR, u) \wedge xabouty \wedge \neg \exists z yaboutz) \wedge \forall y(xabouty \rightarrow \exists z(Concat_3(z, TR, u) \wedge xaboutz \wedge \forall a(zabouta \leftrightarrow a = y)))]$. (infinity)

The above axiom justifies the existence of a propositional function which is about infinitely many propositional functions.

PF10a: $\forall x : AtPropFn(x) \rightarrow PropFn(x)$ (atomic propositional functions are propositional functions)

PF10b: $\forall x : PropFn(x) \rightarrow \exists! y(Concat_3(y, NEG, x) \wedge PropFn(y) \wedge \forall z(yaboutz \leftrightarrow xaboutz))$ (the negation of a propositional function is a propositional function that shares everything that function is about)

PF10c-1: $\forall x \forall y : PropFn(x) \wedge PropFn(y) \wedge \forall a(xabouta \leftrightarrow yabouta) \rightarrow \exists! z(Concat_4(z, CON, x, y) \wedge PropFn(z) \wedge \forall a(zabouta \leftrightarrow xabouta))$ (given two propositional functions that are about the same things, their conjunction is another propositional function about the same things)

...

PF10d: $\forall x \forall y : Var(x) \wedge PropFn(y) \rightarrow \exists! z(Concat_4(z, UNIV, x, y) \wedge PropFn(z) \wedge \forall a(zabouta \leftrightarrow youbouta))$ (given a propositional function with at least one free variable x, the universal quantification of the propositional function over that x is also a propositional function)

From PF10 and PF8, we can derive axiom schemes for propositional functions in general analogous to separation, pairing, union, power set, choice, and replacement.

PF11: $\forall x : Concept(x) \vee PropFn(x)$ (everything is either a concept or a propositional function)

PF12: $\forall x : AtPropFn(x) \rightarrow \phi[x] \wedge \forall x : \phi[x] \rightarrow \forall y(Concat_4(y, NEG, x) \rightarrow \phi[y]) \wedge \forall x \forall y : \phi[x] \wedge \phi[y] \rightarrow \forall z(Concat_4(z, CON, x, y) \rightarrow \phi[z]). \wedge ... \wedge \forall x \forall y : Var(x) \wedge \phi[y] \rightarrow \forall z(Concat_4(z, UNIV, x, y) \rightarrow \phi[z]) \wedge \forall x \forall y : Var(x) \wedge \phi[y] \rightarrow \forall z(Concat_4(z, EXISTS, x, y) \rightarrow \phi[z])) \rightarrow \forall x(PropFn(x) \rightarrow \phi[x])$ (induction scheme for propositional functions)

PF13: $\exists y(AtPropFn(y) \wedge \forall x(youboutx \leftrightarrow Concept(x)))$ (some atomic propositional functions are only about all of the concepts)

Here are the satisfaction axioms. Let z be a propositional function, Q_z be the set of valuation assignments for z, and let s be any variable assignment in Q_z:

S1-1: if z is $IDuv$ then $Sat(s, z) \leftrightarrow s(u) = s(v)$ (satisfaction for identity)

S1-2: if z is $CONCAT_3uvw$ then $Sat(s, z) \leftrightarrow Concat3(s(u), s(v), s(w))$ (satisfaction for two-element concatenation)

S1-3: if z is $CONCAT_4uvwx$ then $Sat(s, z) \leftrightarrow Concat4(s(u), s(v), s(w), s(x))$ (satisfaction for three-element concatenation)

S1-4: if z is $ABOUTuv$ then $Sat(s, z) \leftrightarrow s(u) abouts(v)$ (satisfaction for the about relation)

S1-5: if z is $SATuv$ then $Sat(s, z) \leftrightarrow Sat(s(u), s(v)) \wedge Q_{s(v)} \neq \varnothing$ (satisfaction for Sat)

S1-6: if z is TRu then $Sat(s,z) \leftrightarrow Tr(s(u)) \wedge Q_{s(u)} \neq \varnothing$ (satisfaction for truth)

S1-7: if z is $PROPFNu$ then $Sat(s,z) \leftrightarrow PropFn(s(u))$ (satisfaction for "is a propositional function")

S1-8: if z is $VARu$ then $Sat(s,z) \leftrightarrow Var(s(u))$ (satisfaction for variables)

S1-9: if z is $NEGx$ then $Sat(s,z) \leftrightarrow \neg Sat(s,x)$ (satisfaction for negation)

S1-10: if z is $UNIVux$ then $Sat(s,z) \leftrightarrow \forall s' \in Q_z(s' isu - alternativeofs \rightarrow Sat(s',x))$ (satisfaction for quantifiers - universal case)

S1-11: if z is $UNIVux$ then $Sat(s,z) \leftrightarrow \exists s' \in Q_z(s' isu - alternativeofs \wedge Sat(s',x))$ (satisfaction for quantifiers - existential case)

S1-12: if z is $CONxy$ then $Sat(s,z) \leftrightarrow Sat(s,x) \wedge Sat(s,y)$ (satisfaction for conjunction)

...

S2: $Tr(x) \leftrightarrow \forall s \in Q_x : Sat(s,x)$ (definition of truth for propositional functions)

S3: for all propositional formulas with conceptual variables $X_1, ...X_N$, $Ext(y,x) \leftrightarrow Set_{PF}(y) \wedge y = \{< x1, ..., xn > | \exists s \in Q_x : s(X_1) = x_1, ... s(X_n) = x_n \wedge Sat(s,x)\}$ (extensions of propositional functions)

7 Non-compositional truth theories - The Revision Theory of Truth

In this section we discuss a particular family of instances of truth theories that are not compositional, namely those that fit under *The Revision Theory of Truth*. Here, in general we have the inner theory and truth theory coinciding and being both classical, but compositionality being sacrificed. This is not actually the case for the axiomatic *Friedman-Sheard* revision theory of truth, but since it violates C4, we shall refrain from discussing it. The other obvious candidate for a classical, non-compositional truth theory is Kripke's theory of truth with respect to some supervaluation scheme. However this does not provide us with the notion of *stability* that the revision theory gives us, and it is not clear which valuation scheme to choose from.

7.1 Introduction and summary

The *revision theory of truth* is the name given to several different theories of truth that emerged during the 1980s from writers such as Anil Gupta, Nuel Belnap, Hans G. Herzberger, and Aladdin Yaqub. Each of them draw from Kripke's theory of truth. We shall first talk about the general notion of a "revision sequence" that they each have in common.

Let L^- be a first-order language without a truth predicate, and let M^- be a model for this (i.e. a *ground model*); have L be the result of enriching L^- with a truth predicate. We define a *hypothesis* to be a function h from the set of first-order formulas to truth values, and a *hypothetical valuation* $Val_{M+h}(\phi)$ of a sentence ϕ with respect to a hypothesis h on a ground model M to be the (classical) truth value of the sentence ϕ in $M + h$. First, we suggest an expansion of L^- with a hypothesis. A particularly straightforward example of this is used in [35]: stipulate that all sentences with a truth predicate are false, as given by the hypothesis h_0. Then we change the hypotheses using a *rule of revision* τ_M, which for some hypothesis h evaluating an object d returns 1 if d is the name of a sentence and $Val_{M+h}(\phi) = 1$, and returns 0 otherwise.

For a sequence S of hypotheses of ordinal length γ, write S_α for the hypothesis that is the ordinal α'th element of the sequence. Now say that a sentence d is *stably true* (false) with respect to a sequence of hypotheses S if and only if there exists some ordinal α such that for every ordinal

β such that $\alpha \leq \beta < \gamma$, we have $S_\beta(d) = 1$ (0). With M^- the model of the previous paragraph, an On-long sequence of hypotheses is said to be a *revision sequence* for that model whenever $S_{\beta+1} = \tau_{M^-}(S_\beta)$ for each ordinal β and for each limit ordinal γ and sentence d, if d is stably true (false) with respect to the sequence of the first β hypotheses of s, then $s_\beta(d) = 1$ (0).

Another notion of truth was devised in a similar but independent way by Herzberger in [35], and highlights similarities between the revision theory of truth and Kripke's theory of truth. The difference between Herzberger's theory of truth and Gupta's is purely in the definition of the truth predicate; both use a rule of revision in the same way. Herzberger starts with the hypothetical valuation Val_{M+h_0} where h_0 sends every sentence with an instance of T to 0. The rule of revision may be thought of as providing successor stages in a set-theoretic hierarchy; the limit stage is provided by a *lower limit* operation. That is, for λ a limit ordinal, $S_\lambda(d) = 1$ whenever $d \in llim S_\lambda = \{x : \exists\delta : \delta < \lambda \wedge (\forall\gamma : \delta \leq \gamma < \lambda \wedge (S_\lambda(x) = 1\}$; $S_\lambda(d) = 0$ otherwise. The notions of stability in Gupta and Herzberger coincide; think of a sentence as *unstable* with respect to an initial hypothesis simply if it is not stable given that initial hypothesis. It can be shown that there is a limit ordinal Σ such that a sentence is stably true or false if and only if $T(x) \vee T(\neg x)$ is true at stage Σ; all unstable sentences and their negations are false, and the biconditional T-schema is true for precisely the set of stable sentences. The statements declared true (false) at stage Σ are said here to be *Herzberger true (false).*

With respect to either stable truth or Herzberger truth, the liar is diagnosed as being unstable in every revision sequence for M, the truth-teller is diagnosed as being stably true in some revision sequences and stably false in others, and ungrounded true (false) statements come out as true (false) in all revision sequences. The point of departure for revision theories comes from actually defining truth in these terms. In Herzberger's treatment, the truth and falsity of sentences is declared at a limit ordinal stage in which all unstable sentences are conveniently false - and in general for unstable sentences, and in particular for the liar sentence l, $T('l') \leftrightarrow l$ is false. In any case, the revision theory differs from the accounts of Priest and Field who have the T-schema equivalences holding in the form of biconditionals in the object language.

7.2 Recurring and stable truth

We reflect on the criterion D1. Contingently paradoxical sentences are appropriately diagnosed as being contingently unstable. However, morals can be drawn from brief retrospective accounts of revision theories such as [3] that without some justification, the differing definitions of truth have arbitrary consequences.

The revision theory of truth has the necessary resources to identify paradoxical situations whenever they should arise, based on the pattern of the revision sequences. Truth-teller-like sentences can be distinguished by their being stably true in some revision sequences, and stably false in others. Difficulties begin to arise when drawing the line on what is to be considered true or false, by defining a truth predicate on revision-theoretic terms. Within the paradigm of [28], two possibilities have been considered, though even more, such as Herzberger truth, exist:

1. Recurring truth: let an interpretation J of a set of sentences (with respect to a ground model M^- and a hypothesis h) be called *recurring* if there is some interpretation I with respect to which there are infinitely many n for which $\tau^n_{M^-}[I] = J$. A sentence ϕ of a set of sentences Σ is said to be *recurringly true (false)* if for every recurring interpretation of sentences in Σ with respect to M and h, $Val_{M+h}(\phi) = 1$ (0).

2. Stable truth: let an interpretation J of a set of sentences be called *stable* if for this interpretation, every sentence is either stably true or stably false. A sentence ϕ of a set of sentences Σ is said to be *stably true (false)* if for every stable interpretation of sentences in Σ with respect to M and h, $Val_{M+h}(\phi) = 1$ (0).

Gupta and Belnap have expressed sympathy for recurring truth as at least providing a starting point for a definition of truth; stability corresponds to their theory of definitions S_0, which they consider too weak to be viable.[9] On the other hand, there is a peculiar consequence of the recurring

[9]The example they use to justify this is that in S_0 not all points in a totally-ordered discrete set with a zero element need have the property "either if all of this predecessors have this property then the point itself has this property, or all of this point's predecessors have this property", yet intuitively it must because otherwise there would exist a unique least point without this property, but since all of its predecessors have the property then it must have the property as well.

truth condition raised in [24]. Take the *hemi-tautology*, defined by the following two sentences:

At least one of the next sentence and this sentence is false. Both the previous sentence and this sentence are false.

If the first sentence were false, then the second sentence would be paradoxical. However, if the first sentence were true, then the second sentence would be simply false. So in an "intuitive" proof by contradiction, the first sentence is proved true and the second is proved false. This is not the case, however, if we take truth to be recurring truth: the interpretation where both sentences are hypothesised to be true and the interpretation where both are hypothesised to be false are also recurring, and so it is not the case that the first sentence is recurringly true, and it is not the case that the second is recurringly false.

Gupta and Belnap themselves introduce a new notion of a *fully varied* revision sequence that provides the intuitive solution, but the only motivation for introducing this is to address an example similar to the hemi-tautology above. So while the revision theory provides a good diagnosis of the paradoxical (naively unstable) sentences, the writers resort to ad hoc solutions when actually determining what the true and false sentences are in more obscure cases.

7.3 The perils of the determinacy predicate

The revision theory of truth involves quantification over two-valued interpretations rather than three-valued interpretations, so there is no scope for introducing a novel predicate "this statement is nontrue" or "this statement is only false" as the revenge arguments on the three-valued theories tend to work. But there is no *determinacy* predicate reflecting on how the liar has been diagnosed compared to strengthened liars.

The determinacy predicate can be introduced in the Herzberger theory as an operator D_h on names of sentences, defined as follows:

$D_h('A') = A \wedge T('A')$; $D_h^{\sigma+1}('A') = D_h(D_h^{\sigma}('A'))$; $D_h^{\lambda}('A') \equiv \forall\sigma < \lambda : (TD_h^{\sigma}('A'))$ for λ a limit ordinal

We have another hierarchy of liars, defined as follows:

$Q_0 = l$ (l is the liar sentence); $Q_{\sigma+1} = l*$ where $l* : \neg D_h^{sigma+1}('T('l*')')$; $Q_\lambda \equiv l**$ for λ a limit ordinal, where $l** : \neg D_h^\lambda('T('l**')')$

Philip Welch, Hannes Leitgeb, and Leon Horsten discovered in [26] that having introduced this operator into the language, there must exist sentences that escape the hierarchy of iterations. There is no explicit definition for these sentences, but they can be shown to exist by a proof by contradiction.

Before discussing Welch, Leitgeb, and Horsten's result, some preliminary details should at least be mentioned. Note that for each stable sentence there is a unique ordinal stage in the Herzberger hierarchy with respect to it has a fixed truth value. This ordinal stage can be identified as a predicate ρ acting on names of sentences: $\rho('x') = \alpha$ if and only if the truth value of x is fixed from stage α. (Have $\rho('x') =\uparrow$ if $'x'$ is not the name of a stable sentence) Moreover, in L_T we can set a prewellordering $<'$ of names of stable sentences: $P_{<'}('x','y')$ if and only if $\rho('x') < \rho('y')$. With this knowledge, we may now define for each sentence C an internal hierarchy of determinacy predicates of length $\rho('C')$: $D_h^C('A') \equiv \forall B : P_{<'}(B,C) \rightarrow (\forall y : (y =' D_h^B('A'^{\frown}'0' \rightarrow T('y')))$. Analogously, $Q_C = m$ where $m : D_h^C('T('m')')$.

But now we have:

Theorem: There exist sentences $C \in L^+$ such that with respect to any initial hypothesis g, for any predicate D_h^B with B stable given g, $D_h^B(Q_C)$ is unstable given g.

As mentioned before, Field's process of evaluating the truth of the conditional, a variant of Herzberger's revision process which by [29] returns the same set of true and false sentences as in Herzberger's theory, also falls into the same problems. The substance of the result is that neither Field's theory nor Herzberger's revision theory have a means of diagnosing strengthened liars with determinacy predicates.

7.4 The lesson of revision theories

The various different ways we have seen for the revision process to provide us with a truth predicate (i.e. that of stable truth, Herzberger truth, re-

curring truth) are all in some way questionable. Herzberger's theory comes with no serious justification for why the liar sentence is false (and also unstable), for instance. And as we have seen, there is no justification in the Gulnap/Belnap approach for favouring recurring truth over any other notion apart from getting the desired results out of particular examples. These arbitrary diagnoses predictably lead us to arbitrary results. Just one of these is the hemi-tautology's diagnosis in the Gupta/Belnap theory mentioned earlier.

The revision theories of truth are broadly distinguished by differing *limit rules*; the problem lies in having to choose between them, and the seeming arbitrariness calls into question how useful they would be in a scientific enterprise such as A1 and A2. There is another problem with each of these rules; the complexity of the set of stably or recurring true sentences is high enough to be both implausible from a cognitivist standpoint and unwieldy from an instrumentalist one, as seen in [1].

More than that, it has been remarked in [2] that questions can be asked about sets of sentences that are stably true whose answers are independent of ZFC. The projects of A1 and A2 cannot be advanced in the (stable truth) revision theory of truth without making prior commitments to the foundations of mathematics. Taking an agnostic stance on statements like J, the relative usefulness of certain revision theories in pursuing A1 and A2 can be called into question.

In light of the above, it is apparent that for the purposes of A1 and A2 the question of which limit rule to choose should be set aside in favour of how Gupta's theory of definitions can be applied to another truth theory in a more tentative way, as suggested in [3]. Sentences can be classified as stable, variably unstable, and naively unstable without having to resort to a limit rule such as Herzberger's.

8 Later directions

In this section, the truth theories mentioned in sections four through seven will be assigned roles according to the goals A1-A4 outlined in section two. The roles will be assigned according to how the truth theories cope with the criteria D1-D3 outlined in section three. Proposals for future work will then be given for these truth theories according to the roles they have been assigned. First some words will be given to motivate distinguishing two kinds of truth theory: classical truth theories that restrict the T-schema and transparency, and non-classical truth theories that do not restrict the T-schema or transparency.

A clear distinction between classical and non-classical truth theories was drawn by Vann McGee in [10], in a review of Field's truth theory:

"We have a choice. We can allow ourselves full classical logic and restrict transparency... Or we can uphold full transparency and restrict the logical rules that don't involve "T", so that we are only allowed the full range of classical inferences when the sentences involved don't contain "T"... The classical option has the merits of simplicity and familiarity."

The contrast that can be drawn here is between Field and Leitgeb, transparency (of the truth predicate) and familiarity (of the logic of the truth theory). In restricting the T-schema and transparency, Leitgeb's theory is already ruled out of serving the purposes of A3 or A4. In avoiding traditional revenge issues, Leitgeb's theory serves as a potential foundation for semantics, a possible aid to the cause of A1 or A2.

8.1 A1, A2, and building on L-W

Recall from our earlier analysis of Leitgeb-Welch's truth theory that our objections mostly focused on the conditions for sentences to express propositions:

On D1 - there is no way to distinguish truth teller sentences from liar sentences in the formal theory - both are ungrounded, and thus don't express propositions (H1). Some sentences which based on certain empirical circumstances are ungrounded true in (say) Kripke's theory of truth under the Strong Kleene scheme do not express propositions in L-W (H2). The

condition that two propositions that are both true in L-W must nonetheless be about the same things for their conjunction or disjunction to be true is unusually strong (H3). The disjunction of a true sentence with a sentence that does not express a proposition - and the conjunction of a false sentence with a sentence that does not express a proposition - does not express a proposition (H4).

On D2 - there are no traditional revenge worries for Leitgeb's truth theory. Traditional revenge liar sentences are dealt with in the same way as liar sentences: as counterexamples to the T-schema.

On D3 - the idea that liar sentences do not express propositions, and the idea that some sentences contingently do not express propositions because they are contingently truth-paradoxical, is justified with recourse to contextualist views of meaning and externalist views of cognition. The demand for sentences that express propositions to be grounded is justified with recourse to an analogy to ZF set theory.

Clearly L-W is unfit for the aims of A3 and A4, falsifying both the T-schema and unrestricted comprehension. Leitgeb and Welch explicitly regard their theory as providing a theoretical foundation for semantics, perhaps suiting either A1 or A2. The objections H1-H4 are still pertinent, though, and should be the focus of future work.

The objection H3 may hopefully be circumvented by a strengthening of the relevant axioms of propositional functions, which Leitgeb and Welch remark on speculatively in a footnote. The other objection H1 could be addressed with a more elaborate theory enhanced by revision-theoretic notions. The revision theory in particular provides a means of establishing whether a sentence is paradoxical (no stable interpretations) or truth-teller like (true in some stable interpretations, false in others); additional axioms may be provided determining whether a formula (not expressing a proposition) is paradoxical like the liar or more akin to the truth-teller. H2 and H4 demand either a rebuttal or a more relaxed definition of semantic expression; we draw the line here.

8.2 A3, A4, and Field/Priest - duality

Recall from our earlier analysis of Field's truth theory that there were many objections, mostly focused on Field's new form of implication:

On D1 - by Kripkean methods, grounded true/false sentences can be distinguished from truthteller-like and paradoxical sentences. Paradoxical sentences are treated as being neither true nor false, i.e. as instances of violations of excluded middle, but their being neither true nor false is not expressible in the language. Instead they are, presumably, seen as not determinately true, or otherwise given some place in Field's determinacy hierarchy. But it transpires that this is not even the case for certain sporadic paradoxical sentences (E).

On D2 - a prime motivation for introducing the determinacy hierarchy into Field's truth theory is to evade attacks by means of introducing traditional revenge liars. Traditional revenge liars, in fact, still exist, and must exist in a theory like Field's (F).

On D3 - there is no obvious justification for Field's implication violating the deduction theorem when excluded middle is not satisfied, apart from it being theoretically preferable to violating modus ponens (G1). Field's new form of implication also lacks a scientific basis, again being introduced to provide satisfying theoretical ends (G2). This form of implication is also unwieldy - in particular, so complicated that Field's logic is not even axiomatisable (G3).

Field has already made significant progress in devising a new truth theory with a different, stronger implication operator to supercede the one reviewed here. He has also provided independent justification for accepting paracomplete truth theories. Thus we will not focus here on objection G3, though it would be interesting to see how complicated the new form of implication turns out to be and how intuitive the changes to the truth theory will be. The other objections apart from F are also on shakier ground, though skepticism should be maintained as to whether sporadic sentences cannot exist and new revenge liars cannot be introduced in a prospective paracomplete truth theory. There is also no indication of a change in Field's apparent methodology of tuning definitions of logical operations to satisfy logical rules, or non-ad hoc justification for prioritising the deduction theorem ahead of modus ponens.

With the above in mind, what paradigms A1-A4 would Field's theory fit? A1 and A2 may not be suitable candidates. First, whether sporadic liar sentences are meaningful and in what sense they express a proposition is not provided in Field's theory. Second, notions such as the "bivalent liar" are coherent and expressible in theories such as Leitgeb-Welch's; that they cannot be introduced in Field's still seems to show an explanatory inadequacy. The elephant in the room is that non-sentences are automatically taken to the anti-extension of the truth predicate, as the theory is founded on Kripkean fixed points. Regardless of the non-sentences, it is doubtful whether a range of "revenge" notions can be provided a semantic account by Field's truth theory.

What's more, providing a foundation for semantics does not appear to be Field's goal, though sometimes Field's truth theory is dressed up as providing a "naive" account of truth. Field has expressed skepticism at the coherence of any prospective descriptive theory of truth, considering it to instead be trivial. And we have seen that there is no apparent prospect of a scientific justification of Field's implication or other notions. So much for A1-A2.

With regards to A4, Field's theory provides a support for at least the coherence of a deflationary notion of truth. What's more, the conservativity and strong intersubstitutivity results make it ideal in serving Quine's role for a truth predicate in a logical language as a device of generalisation. We cannot say anything concrete about the prospects for a variant of Field's truth theory being relevant to A3 within these margins, but we are at least optimistic: the truth theory is similar in some ways to the continuum-valued logic of Lukasiewicz, which has been shown to admit unrestricted comprehension without triviality in [6].

Does the above say anything for paracomplete truth theories as supporting the aims A3 and A4, as opposed to paraconsistent ones? We shall see that this is not the case. Recall our discussion of Priest: a truth predicate can be defined for Priest's logic of paradox along dual lines as those for Kripke's theory of truth under the Strong Kleene valuation scheme. It has the same revenge worries as Kripke's theory of truth. There is no obviously compelling justification in principle for accepting a paraconsistent

theory of truth over a paracomplete theory of truth.[10] What we shall find is that there is no obviously compelling justification in principle for accepting a paracomplete theory of truth over a paraconsistent theory of truth, either. The new form of implication, with its strong intersubstitutivity result and determinacy hierarchy, can be carried over to a paraconsistent logic.

All that needs to differ from Field's logic is a "dual" notion of entailment and fixed point construction. This provides us with different results for the implication operator: while the deduction theorem now holds, modus ponens is no longer the case, by the same counterexamples as those Field gives for the deduction theorem.

There may be some cause to advocate an eclectic approach to non-classical truth theories in pursuing the goals A3 or A4. On the one hand we can reject both the liar and its negation, on the other hand we can accept both. We may maneuver around the objection G1 merely by advocating the above eclectic approach to non-classical truth theories.

Revenge issues may be sidestepped by adopting a non-classical model theory or property/set theory, an approach we have purposefully shunned up until now. Nonetheless there have been proposals for such a thing for Field's theory in [11], and they would still serve the logical revisionist aims of A3 and A4 by a methodology similar to that of Bueno and Colyvan (themselves disputing the law of non-contradiction) in [9], adapting Laudan's model of scientific theory change to the choice of logical principles.

8.3 Final remarks

It may now be worthwhile to take stock in what has been written so far. In the second section we found possible reasons A1-A4 for why one would be interested in truth theories and some form of the T-schema, and why the liar poses a special threat in any case. There are many truth theories out there, so we found criteria C1-C8 and D1-D3 by which to categorise and criticise any truth theory. Four kinds of truth theories were chosen on the basis of C1-C8, and from each a suitable candidate was evaluated on the basis of D1-D3. Earlier in this section we reflected on the objections one

[10] Some justification has been appealed to for either paraconsistent and paracomplete theories - one for paraconsistent theories being that the liar is *prima facie* both true and false - but no obvious way, given the arguments, to decide which is more compelling.

could make from an evaluation on D1-D3.

The ultimate aim of the thesis has been to distinguish which truth theories are most successful at solving particular problems or otherwise suiting particular theoretical paradigms. In all cases, the work that is necessary to adequately satisfy the aims of any one of A1-A4 is far from done. Rather than inspiring new work on creating ad hoc means of avoiding the liar paradox in a formal theory, the thesis suggests that existing theories can be modified to suit some of these four aims. The proposals of an eclectic approach to non-classical truth theories and of an augmenting the L-W theory are but two small steps in this direction.

References

[1] Welch, Philip, "On Revision Operators", *The Journal of Symbolic Logic*, Vol. 68, No. 2, June 2003.

[2] Löwe, Benedikt, and Welch, Philip, "Set-Theoretic Absoluteness and the Revision Theory of Truth", *Studia Logica*, Vol. 68, No. 1, pp.21-41, June 2001.

[3] Löwe, Benedikt, "Revision Forever!", *Proceedings of the 14th International Conference on Conceptual Structures*, 2006.

[4] Friedman, Harvey, and Sheard, Michael, "An Axiomatic Approach to Self-Referential Truth", *Annals of Pure and Applied Logic*, Vol. 33, pp. 1-21, 1987.

[5] Hanna, Robert, *Rationality and logic*, MIT Press, 2006.

[6] White, Richard B. "The consistency of the axiom of comprehension in the infinite-valued predicate logic of Lukasiewicz", *The Journal of Philosophical Logic*, Vol. 8, No. 1, pp. 509-534, 1979.

[7] Lukasiewicz, Jan, and Tarski, Alfred, "Investigations into the Sentential Calculus", *Logic, Semantics, Metamathematics* (Alfred Tarski), 1956.

[8] Restall, Greg, "Arithmetic and Truth in Lukasiewiczs Innitely Valued Logic" *Logique et Analyse*, Vol. 139-140, pp. 303-312, 1992.

[9] Bueno, Otávio and Colyvan, Mark, "Logical Non-Apriorism and the 'Law' of Non-Contradiction", *The Law of Non-Contradiction*, pp.156-175, Graham Priest, J.C. Beall, and Bradley Armour-Garb (eds.), Oxford University Press, 2004.

[10] McGee, Vann, "Field's logic of truth", *Journal of Philosophical Studies*, Vol. 147, pp. 421-432, 2010.

[11] Leitgeb, Hannes, "On the Metatheory of Field's 'Solving the Paradoxes, Escaping Revenge'", *Deflationism and Paradox*, J.C. Beall and B. Armour-Garb (eds.), Oxford University Press, 2005.

[12] Welch, Philip, "Some Observations on Truth Hierarchies", Unpublished manuscript.

[13] Welch, Philip, "Ultimate truth vis à vis stable truth", *Review of Symbolic Logic*, Vol. 1, No. 1, pp. 126-142, 2008.

[14] Leitgeb, Hannes, "What theories of truth should be like (but cannot be)", *Philosophy Compass*, 2/2, pp. 276-290, 2007.

[15] Klagge, James, "Convention T Regained" *Philosophical Studies*, Vol. 32, No. 4, pp. 377-381, 1977.

[16] Scharp, Kevin, *Replacing Truth*, Version 3.5, http://people.cohums.ohio-state.edu/scharp1 , 2010.

[17] Grover, Dorothy, "How Significant Is the Liar?" *Deflationism and Paradox*, J.C. Beall and B. Armour-Garb (eds.), Oxford University Press, 2005.

[18] McGee, Vann, "How truthlike can a predicate be? A negative result", *The Journal of Philosophical Logic*, Vol. 14, No. 4., pp. 399-410, 1985.

[19] Van Orman Quine, Willard, "The Ways of Paradox", *The Ways of Paradox and Other Essays*, Random House, 1966.

[20] Gödel, Kurt, "On formally undecidable propositions of Principia Mathematica and related systems I", *Kurt Gödel Collected works, Vol. I*, Solomon Feferman (ed.), Oxford University Press, 1986.

[21] Jech, Thomas, *Set Theory*, 3rd edition, Springer, 2006.

[22] Chang, C.C., and Keisler, H.J., *Model Theory*, 3rd edition, North Holland, 1990.

[23] Mendelson, Elliott, *Introduction to Mathematical Logic*, 4th edition, Chapman & Hall/CRC, 1997.

[24] Cook, Roy T., "Counterintuitive consequences of the revision theory of truth", *Analysis* Vol. 62, pp.16-22, 2002.

[25] Leitgeb, Hannes, and Welch, Philip, "A Theory of Propositional Functions and Truth", Unfinished manuscript.

[26] Horsten, Leon, Leitgeb, Hannes, and Welch, Philip, "Understanding the Revision Theory of Truth", Unfinished manuscript.

[27] Leitgeb, Hannes, "What Truth Depends On", *The Journal of Philosophical Logic*, Vol. 34, pp. 155192, 2005

[28] Gupta, Anil, and Belnap, Nuel, *The Revision Theory of Truth*, 1993

[29] Welch, Philip, "On Revision Operators", *Journal of Symbolic Logic* Vol. 68, pp. 689-711, 2003

[30] Asenjo, Florencio Gonzalez, "A Calculus for Antinomies", *Notre Dame Journal of Formal Logic*, Vol. 16, No. 1, pp. 103-105, 1966.

[31] Priest, Graham, "What's So Bad About Contradictions?", *The Law of Non-Contradiction*, pp. 23-40, Graham Priest, J.C. Beall, and Bradley Armour-Garb (eds.), Oxford University Press, 2004.

[32] Field, Hartry, "Replies to Commentators on *Saving Truth From Paradox*", *Journal of Philosophical Studies*, Vol. 147, pp. 457-470, 2010.

[33] Shapiro, Stewart, "So truth is safe from paradox: now what?", *Journal of Philosophical Studies*, Vol. 147, pp. 445-455, 2010.

[34] Priest, Graham, "Hyper-contradictions", *Logic et Analyse*, Vol. 27, pp. 237-243, 1984.

[35] Herzberger, Hans, "Notes on Naive Semantics", *Journal of Philosophical Logic*, Vol. 11, No. 1, pp. 61-102, 1982.

[36] Gupta, Anil, "Truth and Paradox", *Journal of Philosophical Logic*, Vol. 11, No. 1, pp. 1-60, 1982.

[37] Priest, Graham, "The Logic of Paradox", *Journal of Philosophical Logic* Vol. 8, pp. 219-241, 1979.

[38] Priest, Graham, "Paraconsistent Logic", *Handbook of Philosophical Logic* (second edition) pp. 287-393, D. Gabbay and F. Guenther (eds.), 2002.

[39] Field, Hartry, "Saving the Truth Schema from Paradox", *Journal of Philosophical Logic* Vol. 31, No. 1, pp. 1-27, 2002.

[40] Field, Hartry, "A Revenge-Immune Solution to the Semantic Paradoxes", *Journal of Philosophical Logic* Vol. 32, No. 2, pp. 139-177, 2003.

[41] Field, Hartry, *Saving Truth from Paradox*, Oxford University Press, 2008.

[42] Davidson, Donald, "Truth and Meaning" *Synthse* Vol. 17, pp. 304-323, D. Reidel Publishing Company, Dordrecht-Holland, 1967.

[43] *Deflationism and Paradox*, J.C. Bealle and B. Armour-Garb (eds.), Oxford University Press, 2005.

[44] Williamson, Timothy, and Andjelkovic, Miroslava, "Truth, Falsity and Borderline Cases", *Philosophical Topics*, Vol. 28, pp. 211-244, 2000.

[45] Hintikka, Jaakko, "A counterexample to Tarski-type truth definitions as applied to natural languages", *Philosophia*, Vol. 5, No. 3, pp. 207-212, July 1975.

[46] Beek, Wouter, "Truth-Theoretic Contextualism: Dissolving the Minimalism/Contextualism Debate", ILLC Scientific Publications, Master of Logic Thesis Series, ISSN: 1387-1951, 2009.

[47] Kripke, Saul, "Outline of a Theory of Truth" *The Journal of Philosophy* Vol. 72, No. 19, pp. 690-716, Journal of Philosophy, Inc., November 1975.

[48] Tarski, Alfred, "The Concept of Truth in Formalized Languages", *Logic, Semantics and Metamathematics*, pp. 152-278, Translated by J.H. Woodger, Oxford at the Clarendon Press, 1956.

[49] Belnap, Nuel, and Gupta, Anil, *The Revision Theory of Truth*, Bradford Books, 1993.

[50] Horwich, Paul, *Truth*, Blackwell, Oxford, 1990.

[51] McGee, Vann, "Maximal Consistent Sets of Instances of Tarski's Schema (T)", *Journal of Philosophical Logic*, Vol. 21, pp. 235-241, Kluwer Academic Publishers, 1992.

[52] *Liars and Heaps*, J.C. Beall (ed.), Oxford University Press, 2003.