Chapter 1
Introduction

Theodora Achourioti, Kentaro Fujimoto, Henri Galinon and José Martínez-Fernández

1.1 Presentation of the Volume

In 2011, and within only a few months, four international conferences on truth were independently organized in Amsterdam ("Truth be told", 23–25 March 2011), Barcelona ("BW7: Paradoxes of truth and denotation", 14–16 June 2011), Paris ("Truth at work", 20–23 June 2011) and Oxford ("Axiomatic theories of truth", 19–20 September 2011). This succession of events and the original work presented at them are evidence that the *philosophy of truth* is a lively and very diverse area of study. They saw a great variety of methodologies from philosophers, logicians and linguists, and even within these groups, a variety of problems and approaches to those problems. We think, however, that the interaction between the different research programmes was not as intense as it could have been. By collecting in one volume a wide range of the very latest research on truth, we hope to intensify the dialogue between philosophers and thus make a contribution to even better informed

T. Achourioti
ILLC & AUC, University of Amsterdam, Science Park 113, 1098 XG,
Amsterdam, The Netherlands
e-mail: t.achourioti@uva.nl

K. Fujimoto
Department of Mathematics and Department of Philosophy,
University of Bristol, University Walk, Clifton, Bristol BS8 1TW, UK
e-mail: kentaro.fujimoto@bristol.ac.uk

H. Galinon
PHIER, Université Blaise Pascal, Clermont Ferrand, France
e-mail: henri.galinon@univ-bpclermont.fr

J. Martínez-Fernández
Lògica, Història i Filosofia de la Ciència, Universitat de Barcelona, Montalegre 4,
08001 Barcelona, Barcelona, Spain
e-mail: jose.martinez@ub.edu

T. Achourioti et al. (Eds.), *Unifying the Philosophy of Truth,* Logic, Epistemology,
and the Unity of Science 36, DOI 10.1007/978-94-017-9673-6_1

research in the future. Hence the title of this volume, *Unifying the philosophy of truth*, which announces our project.

We are very glad that Springer agreed to host this volume within its 'Logic, Epistemology, and the Unity of Science' series. Although—as illustrated by the essays in this volume—contemporary research on truth is now mainly pursued in full independence of the early unity of science programme, this is a good place to recall that positivism played an essential role in the birth of contemporary research on truth in the first half of the twentieth century. The spirit of scientific empiricism and logical rigor promoted by positivism was at first at odds with the use of the concept of truth in philosophy and science. In the 1920s the philosophical notion of truth was under threat from various angles. First, it was not clear at that time what the appropriate conceptual analysis of the notion of truth should be. The traditional conception of truth as correspondence between discourse and what discourse is about was perhaps shared by many philosophers, but it was not seen as amenable to analysis in logical and empirical terms as positivists required for meaningful discourse. Thus, the metaphysical notion of truth was rejected. Second, it seemed all too evident to some philosophers[1] that the rehabilitation of the notion of truth would be a threat to the way they had conceived of their physicalist unitarian project. If the traditional notion of truth were accepted among scientific notions, that would pave the way for truth-conditional semantics, which would then threaten verificationist semantics as the scientific basis for the explanation of meaning. However, although it seems relatively clear that a sentence such as the then all too famous "Das Nichts selbst nichtet"[2] has no verification conditions, it is not equally clear that truth-conditions cannot be used instead to specify its meaning. In other words, the positivist critique of metaphysics would lose its bite if the language of metaphysics could be shown to be meaningful in terms of truth-conditions. Third, even in mathematics, truth-talk was not very fashionable in those days, as Hilbert's formalist programme was very prominent as a philosophy of mathematics.[3] And finally, the concept of truth had notoriously been involved in paradoxes for more than two thousand years. Even if it had not always been clear whether those paradoxes should be taken seriously or not[4], this could not have helped build trust in the notion of truth as a legitimate one, even less so at the beginning of the twentieth century which *was* a time—if any were—in which paradoxes were taken seriously. It is in this historical context that Tarski—following up on work within the Polish school[5]—published his celebrated essay on the notion of truth in formalized languages (Tarski 1983), giving birth to

[1] Neurath in particular. See, e.g. Mancosu (2009).
[2] See, e.g. Carnap (1931).
[3] Remember that in this context Gödel himself, despite his realist convictions in mathematics, carefully avoided use of the term 'truth' in his incompleteness paper. See Feferman (1989).
[4] But see Read's paper in this volume with respect to the Middle Ages.
[5] On the roots of Tarski's work in the Polish school, see Wolenski (2009) and Wolenski and Murawski (2008).

the contemporary research on truth.[6] Part of Tarski's philosophical work and long-standing success in the rehabilitation of the notion of truth is explained by his meeting the positivists' strictures[7] showing, in effect, how in many circumstances the notion of truth can be rigorously defined in a scientifically acceptable language: a paradox- and metaphysics-free language for science. He did so in a series of writings widely acknowledged as a model of balance between philosophical insight and scientific achievement.

Today, another close connection between truth theorizing and the unity of science programme is found in deflationism. Deflationism is probably the most discussed philosophical approach in contemporary philosophical research on truth and it appears in several places in this volume.[8] Some versions of deflationism have been motivated by concerns raised by forms of physicalism or radical empiricism with the nature of truth-theoretical explanations.[9] For it has been argued that physicalism implies that the property of truth is reducible to physical properties or, if not, that it must have no explanatory force. But a full reduction of the property of truth to empirical properties appeared to be hard to achieve[10] and this put some pressure on the physicalist to accept that the notion of truth has no explanatory force. Rather than return to a pre-Tarski state of affairs in which the notion of truth is removed from the language of science, the deflationist's twist is to maintain that the notion of truth is still legitimate in scientific talk, but for logical and not explanatory purposes. This position has in turn stimulated a number of discussions to which philosophy, logic and linguistics have all made contributions, some of which are discussed in detail in this volume.[11]

It seems safe to say that most philosophers, logicians and linguists do not adhere to the early—or late—positivist unity of science programme. But from that tradition we retain the goal of a scientifically informed philosophy of truth. Contemporary philosophy of truth has lain all along at the crossroads of logical, empirical and

[6] One has to add that in 1935 the context had already dramatically changed in the philosophy of mathematics after the publication of Gödel's incompleteness theorems in 1931.

[7] For more on Tarski's relationship with philosophers close to the Vienna Circle, see e.g. Mancosu (2008a, 2008b, 2009).

[8] This is not to say that most philosophers are deflationists. Indeed, Bourget and Chalmers' recent survey of professional philosophers (Bourget and Chalmers 2014), shows that correspondence theory is still the most widely shared view on truth. Within the sample faculty population (admittedly strongly biased towards North American philosophy departments), the results for the different conceptions of truth are as follows ("accept" and "lean towards" answers are aggregated): correspondence 50.8 %; deflationary 24.8 %; epistemic 6.9 %; other 17.5 %.

[9] We have in mind the deflationist tradition running from W. V. Quine (1970, 1960) to H. Field (2001), including the work of, e.g. S. Leeds (1978). P. Horwich (1998a, 1998b) could perhaps also be attached to this tradition, even though he seems to conceive of his deflationism as a philosophical elucidation in the tradition of Wittgenstein's philosophy of language.

[10] Thus, according to Field (1972), Tarski's work did not establish the reducibility of truth to empirical properties. Later attempts, such as e.g. Fodor (1989) in the context of a defence of "intentional realism", have also been criticized. See Loewer (1997) for an overview of naturalizing semantics. Field (2001) illustrates clearly Field's route from physicalism to deflationism about truth.

[11] See in particular Chapts. 4 and 5 of the present volume.

philosophical research, stimulating ever more interaction between them: applications of logical methods help decide philosophical matters (e.g. on the nature of truth[12]), philosophical reflection informs logic research (e.g. on paradoxes[13]) and so on. We hope that the present volume further encourages this ongoing dialogue between philosophy, logical methods and empirical work.

1.2 Organization of the Volume

Research in the philosophy of truth has expanded in many different directions over recent decades. The present volume could not possibly cover the full range of actively researched truth-related topics; however, it does provide an overview of some of the main themes that run through the work currently undertaken within the area in the analytic tradition. This is done directly, through the broad range of topics that the papers address, as well as indirectly, via the authors' reference to others' work that relates to their own.

We have grouped the papers into six chapters (2–7 of this volume). Here we introduce each of the chapters starting with a general presentation of the papers they contain in the order that they occur. This is to give the reader a first idea of what the papers in each chapter are about before we go on to introduce each of them separately. An introduction to each of the papers contained in the chapter then follows.

The topics addressed in different sections of this volume often relate to each other; by no means do we consider our organization of the volume the only possible one. Some of the connections are pointed out by the authors themselves, others we try to highlight in our introduction. A goal that we hope to have achieved by putting the papers side by side the way that we have is to help draw philosophical connections between the papers that go beyond the particular methodologies used. We have, therefore, opted for a divide that cuts across the usual distinctions—distinctions we do not really ascribe to—e.g. between philosophical, logical and linguistic papers. We hope that this, perhaps somewhat unorthodox, organization of the volume will prove helpful in further emphasizing connections between the papers by also drawing the readers' attention to work that they may not at first consider as immediately relevant to their own.

The introduction of the separate papers is not balanced. For each paper we present what we anticipate will be most useful to the non-specialized reader. For example, in cases where we judge that background knowledge of certain issues is required, we have tried to provide part of that background, occasionally at the expense of expanding on the paper's original ideas. In cases where the argumentation of the paper is rather complex, we have opted for a concise presentation of the argumentative structure or an exposition of the preliminary context which then makes it easier to

[12] See Ketland (1999) and Shapiro (1998) on deflationism and conservativity, and Chap. 5 in this volume.

[13] See, in this volume, Chaps. 6 and 7.

penetrate and appreciate the work. With this selective approach we hope to facilitate a comparative reading much more than could be achieved by a uniform exposition, which in some cases would favour the specialized reader.

Finally, it is often said that truth theorists should clarify their philosophical aims and presuppositions before delving into technical details; but we believe that philosophical deliberation and logical analysis should go hand in hand and complement each other. The interaction between ideas and formal techniques is generally a highly complex and intricate affair. Philosophical thinking may help clarify the ideas behind formalization, while reflection on technical work may lead to progress in philosophical thinking. This volume can be read as an illustration of this interactive process.

1.2.1 Truth and Natural Language

The first two papers in this volume offer a critical reflection on the transition from natural to formal language and to philosophical theories of truth. It is common for philosophers to use examples of sentences in natural language that contain the word 'true'. For example, deflationists often explain the non-substantive character of truth by noting that a sentence of the form 'it is true that A' has the same meaning as 'A' itself (an idea that goes back as far as Frege 1918). In her essay, Moltmann studies the linguistics of truth predication which she regards to be a complex phenomenon of natural language. A linguistic study of truth is of immediate relevance not only to philosophers who want to formalize the ordinary notion of truth but also to those who simply wish to align their theory with the behaviour of truth in natural language.[14] Moltmann's analysis shows that far from supporting a single philosophical theory, some of the ways truth predication manifests itself challenge views of truth that are as prominent as deflationism, or those that consider propositions to be the primary bearers of truth.

In a more foundational approach to the very endeavour of formalizing the notion of truth in natural language, Collins objects to a dilemma that seems to drive some of the contemporary discussions: either a consistent theory of untyped truth has to be developed or natural language is inconsistent because it contains a paradoxical notion of truth. Collins' view is that the paradoxical character of truth is inescapable and yet this does not imply that natural language is inconsistent, because questions of consistency can only meaningfully apply to formal theories, which natural language is not. In the essay following Collins', Sheard claims that it is still possible to identify consistent uses of an inconsistent notion and, furthermore, that such consistent uses are evidenced in the case of truth by the fact that paradoxes do not hinder speakers of ordinary language when communicating with each other when using the word 'true'.

[14] Think of questions concerning whether truth should be formalized as a predicate or an operator, whether it is an iterable notion, etc.

These consistent fragments of the use of the truth predicate in natural language can be analysed as inferential mechanisms wedded to specific communicative tasks. One can then study, in a spirit of truth-theoretic pluralism, which of the available axiomatic theories of truth offer the principles needed for carrying out separate tasks; thereby setting specific standards against which existing theories can be adjudicated (which is desirable nowadays given the number of interesting axiomatic theories of truth available).

1.2.1.1 "Truth Predicates in Natural Language" by Friederike Moltmann

Moltmann takes a close look at the appearance of truth in natural language and asks whether the linguistic data support known philosophical views of truth; or weaker than that, whether they are compatible with them. She does not focus her critical study on one particular philosophical theory, nor is there one such theory that is naturally favoured by the linguistic data provided by her study, although some philosophical positions are either excluded or significantly challenged. In fact, this paper is in line with so-called truth-theoretic pluralism: the view that there may be more than one viable notion of truth. Pluralism regarding truth is not a recent view; it has famously been defended, for example, by Lynch (2009). In the present volume, truth-theoretic pluralism is also found to be agreeable by Halbach and Horsten, who take their cue in this from Sheard (1994).

It is an essential assumption underlying Moltmann's analysis that truth predication be regarded as a phenomenon that extends beyond the occurrence of the word 'true'. Truth can be predicated by several expressions, here called 'apparent truth predicates', of which the standard truth predicate is only one. 'Apparent truth predicates' can either ascribe the property of truth (type 1 truth predicates), or express a relation of truth (type 2 truth predicates).

With respect to type 1 predicates, Moltmann argues that the semantics of natural language does not support an operator analysis. Such an analysis for 'is true' has been proposed, for example, by Grover et al. (1975), Grover (1992), Brandom (1994) and Mulligan (2010), and recent formal work on truth has again raised the question of whether truth should be formalized as an operator. Moltmann shows that a truth predicate does not exhibit distinctive sentential semantics such as one finds with expressions that are clear cases of operators, e.g. 'is possible'. She also shows that there is no reason to regard the linguistic form 'it is true that A' (that-clause in extraposition) as more representative of the occurrence of truth in natural language than its equivalent 'that A is true' (that-clause in subject position); whereas an operator approach does favour the former over the latter. Moreover, it is shown that there is no more reason to study constructions in which truth is predicated of that-clauses than nominal expressions; the latter almost universally neglected by philosophical theories of truth.

Under type 1 truth predicates, Moltmann also places normative predicates that are used to convey truth, such as 'is correct' or 'is right'. This gives rise to two notions of truth: representation-related and normative-related, which are combined in normative

truth-predicates and difficult to separate out. Moltmann's proposal is that studying the semantics of these normative predicates will provide insight into the nature of truth predication itself. She concludes, for example, that truth is predicated over intentional entities (attitudinal objects, see Moltmann 2003, 2013) rather than mind-independent entities, such as propositions (or sentences), which is what deflationists about truth traditionally claim (e.g. Horwich 1998). Note that the view that truth is predicated over intentional objects such as beliefs is already found in Ramsey (see Ramsey 1991, p. 8). It follows that the viability of the deflationist view for such predicates, given their semantics, depends on the possibility of distilling a purely representational role for truth.

1.2.1.2 "Truth and Language, Natural and Formal" by John Collins

Collins' essay is as much about the use of truth in natural language as about the paradoxes. For its starting point, recall that Tarski sees the paradoxes as the outcome of (1) the T-biconditionals that characterize the concept of truth, (2) the classical logic that we employ in reasoning about truth and (3) the fact that natural language can speak about everything and in particular it can speak about itself. This diagnosis leads Tarski to what has been called the 'inconsistency view' of truth: since the T-biconditionals are essential to define truth and classical logic should not be modified, one has to admit that paradoxes are produced because natural languages are universal, i.e. they contain their own truth predicate. This implies that natural languages use truth in an inconsistent way. The solution should therefore come from creating non-universal thoroughly-specified formal languages with rich expressive resources that can consistently incorporate a truth definition that implies all the T-biconditionals. And Tarski showed us how to do just that.

The Tarskian analysis of truth has been severely criticized and a new orthodoxy has been developed: the expressiveness of natural language should not be compromised, and the new goal in solving the paradox is to devise powerful formal languages that can speak about themselves to the extent of expressing all paradoxical sentences and still have a consistent truth predicate, even if this implies tinkering with classical logic.

In his paper Collins wants to challenge both sides of this discussion: he wants to defend, against most contemporary solutions to the paradox, the notion that Tarski was right in his diagnosis of the inherently paradoxical nature of the notion of truth in natural language; but he also wants to criticize defenders of an inconsistency view of truth for understanding natural language as inconsistent. Collins proposes a different interpretation of Tarski's view, one that respects the basic intuition that paradoxes are insoluble in natural language, while, at the same time, it does not see natural language as inconsistent.

The first part of the paper compares natural and formal languages. Formal languages are characterized by having an explicit stipulation of their syntax and a full transparent semantics. The full transparency of the semantics means that the syntactic conditions express the semantic properties of the language in such a way that

the semantic properties can be read off from the syntax. Formal languages are guaranteed by design to have these features. In contrast, natural languages are not fully transparent. Collins develops this idea by focusing on five linguistic phenomena: (i) ambiguity, which shows that syntactic structure is not always an accurate guide to interpretation; (ii) the presence of words that do not make any contribution to the sentences they appear in; (iii) the absence of words that should be present in a sentence; (iv) the abundance of positions in sentences that do not serve to predict the interpretation of their occupiers; and (v) the fact that there is no decidable notion of being a well-formed formula of natural language because the acceptance and the meaning of sentences depend on the psychological states of speaker-hearers of the language.

In the second part of the paper, after rejecting two objections to his understanding of natural languages, Collins focuses on the concept of truth in natural and formal languages. Collins claims that the concepts of consistency and inconsistency apply only to formal languages and cannot apply to natural languages because a natural language is not a set of fully transparent sentences such that we could have a consistent or inconsistent theory of it. He criticizes the opposing views of two contemporary defenders of the inconsistency view: Eklund and Patterson (Eklund 2001; Patterson 2008, 2009). Collins also criticizes the views of authors who propose formal languages that could model the universal aspect of natural language and still escape (or at least modulate) inconsistency. Against these views, Collins argues that paradoxical arguments are unavoidable in natural language, due to the inherent riskiness of the truth predicate. This riskiness is produced because, as Kripke (1975) pointed out, one can predicate truth of a set of sentences without knowing the content of the sentences themselves.

1.2.1.3 "Truth and Trustworthiness" by Michael Sheard

Sheard starts by observing that, as opposed to the theoretical case, the use of an untyped truth-predicate in real-life communication appears unproblematic: natural language users have no problem understanding each other when they use the word 'true', irrespective of their potentially different philosophical ideas concerning truth. Sheard proposes this as evidence of some kind of inferential semantics that operates on the surface level of language use and which is shared by language users. This inferential semantics must be consistent, since paradoxes do not seem to arise (or are somehow avoided) in everyday communication. The alternative to allowing for consistent uses of the truth predicate would be to consider language users as irrational beings, in which case empirical psychological work should explain how they manage to deal with inconsistencies; but, according to Sheard, this is a more general question which would not necessarily shed any light on the question: what is the inferential mechanism that is at work in specific situations when people use the word 'true'?

Sheard focuses on the use of 'true' in its function of conveying information. He constructs a simple scenario consisting of two (idealized) agents, a speaker and a hearer, where the speaker conveys a message with the help of the truth predicate by

means of an assertion, a denial or a generalization. The task of the hearer is to decode the speaker's message and to assimilate the knowledge it contains. Sheard then asks of each of three prominent axiomatic theories of truth, so-called FS, KF and VF[15], whether it provides a mechanism for the hearer to perform this decoding act. He observes that this very much depends on whether the hearer considers the speaker to be a trustworthy source, since if not, the hearer first has to check that the message does not lead to inconsistency before assimilating it. Decoding a message is pretty straightforward for all three systems and message forms in the case of a trustworthy source, with the exception of denial, which one has to formalize $T(\ulcorner \neg A \urcorner)$ instead of $\neg T(\ulcorner A \urcorner)$ (this is because decoding $\neg A$ from $\neg T(\ulcorner A \urcorner)$ requires the inference from A to $T(A)$ which is not generally available). Matters become complicated in the case of an untrustworthy source. Logically, KF and VF are equipped to deal with incoming inconsistencies since they are closed under *reductio ad absurdum*, but note that idealization assumptions become crucial here, i.e. the ability of the hearer to screen all logical consequences of existing knowledge for inconsistency.[16]

Decoding messages is discussed by Sheard as a seemingly simple example of a communicative task that allows a comparison between different axiomatic theories of truth. One should not, however, be surprised if a certain theory that fares well in this context does much worse than other theories once one changes the task at hand—Sheard gives another example to this effect. In fact, one should not expect there to be a single axiomatic theory that can account for all communicative uses of truth.

Sheard's approach is, therefore, compatible with both truth-theoretic pluralism and inconsistency theories of truth, since an inconsistent notion may allow for mutually incompatible, yet consistent, uses of it. There are two main reasons for engaging in this exercise of assessing axiomatic truth theories against simple communicative tasks: first, it offers more insight into the philosophical and inferential import of these theories; and second, it provides criteria for adjudicating between the theories. The theme of adjudicating between axiomatic theories of truth is also taken up later in this volume by Halbach and Horsten in chap. 12.

1.2.2 Uses of Truth

Emphasis is also placed on the non-paradoxical features of truth by Rouilhan who demonstrates how Davidsonian truth-theoretic meaning explanations could be used

[15] For an exposition of these theories, the reader can consult Halbach (2010).

[16] Sheard notes that FS presents the additional difficulty of not allowing for *reductio* reasoning due to its lack of a general deduction theorem. The best one can do, Sheard explains, is to provisionally accept a message and use the FS inference rules as a test mechanism in order to track potential inconsistencies while resisting the final step of the reductio argument, which means that instead of adding the negation of the provisionally accepted message to the database, the hearer simply dismisses the message altogether.

to debunk another paradox. 'Frege's paradox'—or rather, a general version of it—arises when meaning explanations for a language intended to be used as the language of science must make use of grammatical categories which clash with the logical structure of that language. Rouilhan's paper shows that appropriate uses of the notion of truth make it possible to give meaning explanations for a language of science that obey its type-theoretical logical structure, and thus comply with the universalism of founders of modern logic[17], that is, in a way that does not condemn these explanations to being nonsensical.

Kahle's 'Sets, Truth, and Recursion' illustrates how the notion of truth can be applied to foundational topics in mathematics, especially set theory. More specifically, Kahle's essay presents a set theory based on an axiomatic truth theory, where sets and the membership relations are *defined in terms of a truth predicate*. The set theory is a so-called *Frege structure*, roughly a way to restrict truth-theoretic assumptions and objects of the theory so as to maintain full comprehension. It is then a consequence of these restrictions that Frege structures support a non-classical concept of truth.

Still on the foundational side, Eberhard and Strahm explore the use of truth in 'unfolding' the content of arithmetic theories. The *unfolding programme* has famously been developed by Feferman and addresses a query of Kreisel's about the proof-theoretic commitments that one implicitly makes when accepting a certain theory. Eberhard and Strahm previously worked on theories of truth for feasible arithmetic, that is, arithmetic weaker than PA, which describes feasibly computable functions (usually identified with polynomial time algorithms). Here they consider the use and strength of such theories in carrying out the unfolding programme.

Finally, Bruni explores a fragment of the revision theory of truth, which was famously developed by Herzberger, Gupta and Belnap as a response to the truth-theoretic paradoxes. In particular, Bruni focuses on the finite use of a technique, the revision rule, which was primarily meant as a way to provide a natural semantics for predicates expressing circular concepts, such as the truth predicate. Finite revision falls short of giving a semantics for the truth predicate, yet it finds natural applications in, for example, game theoretic settings, where players base their decisions on what is rational for other players to do. Besides these applications, Bruni also highlights interesting connections between the finite fragment of revision theory and the FS axiomatic theory of truth.

1.2.2.1 "Putting Davidson's Semantics to Work to Solve Frege's Paradox on Concept and Object" by Philippe de Rouilhan

In his contribution, Rouilhan introduces the reader to what, in the paper itself, he calls for short *Frege's paradox* and the *generalized Frege's paradox*. These paradoxes are not part of the family of Russell's well-known paradoxes that afflicted Frege's logical

[17] See e.g. Rouilhan (2012) and references therein for a broader philosophical perspective on logical universalism.

system, nor part of the family of Frege's puzzle about identity statements; neither are they paradoxes concerning truth. Rather, for once, and as the title suggests, the concept of truth is not part of the problem, but part of the solution.

So, just what is Frege's paradox? It first arose as a consequence of Frege's conception of a language of science. For Frege, as is well known, a predicative term refers to a concept, a singular term refers to an object, and concepts are not objects. But then, to explain the semantics of a language of science, Frege felt that he was inevitably led to say things like, e.g. *the concept horse is not an object*. For this very sentence to have a meaning, according to Frege's logical grammar, the expression *the concept horse* must itself refer to an object. But if *the concept horse* refers to an object, it fails to refer to the concept horse itself (concepts are not objects!)- and so would any expression of the appropriate logical category to serve as subject for the predicate *is not an object*. Thus, what is intended, Frege thought, cannot properly be said, thereby leading to a kind of ineffability thesis.

Rouilhan generalizes Frege's conception of the logical grammar of a language of science. He calls 'generalized Frege's paradox' that which arises whenever one devises a putative language of science such that, in order to explain what this language means, one must resort to another language whose logical grammar clashes with the logical grammar of the language in question. Rouilhan thinks that, despite the fact that Frege himself was prepared to live with the paradox, falling prey to it is really anathema to any putative language of science. So the question is: is it possible to escape the paradox; and if so, what types of language of science do escape it?

Rouilhan argues that the paradox can be escaped for a wide range of plausible candidate languages of science. His starting point is to take up Davidson's idea that explaining what sentences in a language mean is to give their truth-conditions in the form of a 'recursive theory of truth à la Tarski' for that language. One has solved the generalized Frege's paradox for a putative language of science, Rouilhan argues, if one is able to construct a recursive definition of truth for the language in another language that complies with the logical grammar of the language under consideration. Hence the question arises: is it always possible to do so? If the logical basis of the language of science is ZFC, Rouilhan recalls that one can devise a recursive definition of truth for such a language in an extension of the language that shares the same logical basis and only has further extra-logical primitive vocabulary. The meaning-explanation of the language of science is thus carried out in accordance with its logical grammar, without any category mistake with respect to it, or any involvement of entities that were not taken from the start to be part of its ontology. But what about candidates for the status of language of science whose logical basis is type theory or a part of it? Do they escape the generalized Frege's paradox? The answer is far less straightforward and Rouilhan's contribution here is to show that it is still possible to show that they do.

True, it can be shown that one cannot construct a recursive truth definition for a language of infinite order in another language that shares the same logical basis. But Rouilhan submits that there are independent reasons why an infinite order language is a dubious candidate for the status of language of science anyway. The question of interest, then, concerns languages of given finite order. To illustrate, consider a

monadic language, L, of finite order n—that is, with typed variables ranging over individuals or over classes of individuals or over classes of classes of individuals, etc., up to classes of order n, without overlap, and nothing more. Is it possible to construct a recursive definition of truth for L in an extension of it of the same order, obtained by adding, at most, a few extra-logical constants? Rouilhan proves that it is possible if $n \geq 4$. Thus, one can explain the meaning of L in a Davidsonian manner if $n \geq 4$ and, in the end, the generalized Frege's paradox proves to be no threat to adopting such a language as the language of science. Of course, as the author reminds us, recursively defining truth for any language in another language sharing the same logical basis is not the same as giving any definition of truth for the language in question in itself: dealing with this latter difficulty brings us back to the familiar paradoxes of truth, and falls outside the solution of the generalized Frege's paradox.

1.2.2.2 "Sets, Truth, and Recursion" by Reinhard Kahle

In his contribution, Kahle presents a theory of sets by means of an axiomatic theory of truth by defining sets and membership relations in terms of a truth predicate: namely, he identifies an object a being a member of a set $\{x \mid P(x)\}$ for a predicate P, with the predicate P being true of an object a. This idea of Kahle's is based on the notion of so-called *Frege structures*, and he has published a series of papers on theories of Frege structures over *applicative base theories* (Kahle 1999, 2001, 2003, 2009, 2011). The present paper provides an overview of his work and presents the philosophical foundation of his framework.

The concept of Frege structure was introduced by Aczel to "isolate the structure of that part of Frege's *Grundgesetze* that we consider to be correct" (Aczel 1980, p. 38) and he illustrated a semantic construction of a Frege structure over models of lambda calculus. Beeson (1985) gave the first axiomatic system F of a Frege structure. The idea of a Frege structure in connection with the theme of this volume could be summarized as follows.

(1) The full comprehension axiom should hold for all propositional functions: namely, every propositional function f forms a set $\{x \mid fx\}$ such that $a \in \{x \mid fx\}$ is a true proposition iff fa is a true proposition.
(2) Russell's paradox is caused by Frege's assumptions that (i) propositions are either true or false and the conception of the truth of propositions is the classical bivalent one, and that (ii) every formula (or well-formed expression) gives a propositional function.
(3) A Frege structure determines what objects are propositions (and true propositions) and thereby provides the definitions of sets and membership relations so that the full comprehension axiom in the form of (1) above consistently holds by abandoning Frege's two assumptions (i) and (ii).

In the literature, a Frege structure is usually formulated over combinatory algebras (models of combinatory logic) or λ-structures (models of λ-calculus) which are special kinds of applicative structures. Applicative structures are meant to deal with certain abstract conceptions of functions ("functions as rules" in Barendregt 1984 or "functions as operation processes" in Hindley and Seldin 2008), although it is debatable precisely what this means. Each member a of the domain D of an applicative structure can be "applied" to any member $b \in D$ and thereby yields an output $ab \in D$; note, however, that elements of D are not functions in the set-theoretic sense because they are, so to speak, universal functions that can be applied to everything; see Barendregt (1984, Chap. 1) and Hindley and Seldin (2008, Chap. 3E), for further discussion. So, a Frege structure counts propositions and propositional functions among the objects of the domains of applicative structures; this assumption might be justified by arguing that propositional functions are "functions" anyway and propositions are the values of propositional functions. Hence, the intended domain of a Frege structure contains the bearers of truth (*not* sentences *but* propositions in this setting) together with various other mathematical objects and functions. It is usually assumed that propositions are logically suitably structured so that these structures possess some distinguished syntactical operations such as $\dot{\neg}$, $\dot{\wedge}$, etc., whose intended interpretations are the functions that send propositions to their negation, their conjunction, etc.; a discussion concerning the assumption of a logical structure of propositions in the context of formal theory of truth can be found, e.g. in Halbach (2010, § 2).

In the usual presentation of a Frege structure, as offered by Aczel, Beeson, Cantini (whose monograph Cantini (1996) is also an important reference on Frege structures) and Kahle, a set $\{x \mid fx\}$ for a propositional function f is defined as $\lambda x.fx$ by means of λ-abstraction, which is available in combinatory algebra, and a proposition $a \in \{x \mid fx\}$ is simply defined as $(\lambda x.fx)a$. Consequently, since $a \in \{x \mid fx\}$ is just identical to fa by definition (or β-reduction), the full comprehension axiom in the form of (1) above is in fact a trivial and immediate consequence of the definitions of sets and the membership relation $\in$. Hence, in the customary setting, the construction of a Frege structure essentially comes down to the question of how to give a sensible characterization of propositions and truth.

In axiomatizing Aczel's semantic construction of a Frege structure, Kahle adopts an *applicative theory* TON as the base theory of his theory FON of Frege structure. Applicative theories are first-order theories for applicative structures and are usually assumed to include combinatory logic as their core component; for an exposition of combinatory logic, see Barendregt (1984) and Hindley and Seldin (2008), or see Cantini (1996) for its connection to axiomatic truth theories. Then he introduces a truth predicate T as a primitive predicate symbol, and expresses "x is a proposition" by $Tx \vee T\dot{\neg}x$. Thereby, for example, the full comprehension axiom in the form of (1) above can be expressed by:

$$\forall a(Tfa \vee T\dot{\neg}fa) \rightarrow \forall a[T(a \in \{x \mid fx\}) \leftrightarrow T(fa)], \qquad \text{(FCA)}$$

where $\forall a(Tfa \vee T\dot{\neg}fa)$ expresses "the value of f is always a proposition", i.e. "f is a propositional function". As mentioned earlier, FCA trivially obtains by definition,

and the essence of Kahle's axiomatization of Frege structure lies in the postulation of appropriate axioms for T so that it properly expresses truth. However, since Frege's two assumptions must be restricted to sustain consistency, the conception of truth in a Frege structure is inevitably non-classical. In fact, truth in Kahle's theory FON behaves in accordance with the non-classical strong Kleene logic; namely, the "inner logic" of FON is strong Kleene logic. As another example, the inner logic of the truth in Aczel's original semantic construction and Beeson's axiomatization F is non-classical Aczel-Feferman logic (see Fujimoto 2010, where it is called Feferman Logic).[18]

Lastly, let us explain a notable difference between Kahle's FON (as well as Eberhard and Strahm's theories introduced later on) and the more traditional type of axiomatic theories of truth such as Leigh's and Cieslinski's in this volume. In the traditional setting, a truth predicate is conceived of as a predicate of sentences (i.e. sentences are taken to be the bearers of truth), and syntactical objects such as sentences are considered to be distinct objects from those of the subject matter of theories of truth (such as natural numbers and sets). Accordingly, a base theory B such as PA of the traditional type of axiomatic theories of truth has to perform two totally different roles at the same time (i.e. that of a theory of the subject matter and that of a theory of syntactical objects *via a certain coding system* such as Gödel numbering) and two totally different types of objects (i.e. the mathematical objects of the subject matter and syntactical objects) are entangled in the domain of discourse of B. This entanglement causes, for example, the following problem of axiomatic schemata: when a truth predicate is newly introduced into an arithmetical base theory B, one might want to expand the arithmetical induction schema for the augmented language so as to enable arguments or proofs by induction on the syntactical complexity of sentences, on the one hand; but one might also want not to expand the induction schema for the augmented language so as not to make further mathematical commitments from a deflationist point of view, on the other hand. In other words, it might be the case that one wants to expand the schema in regarding B as a theory of syntax but also wants to restrict the schema in regarding B as a theory of mathematics. For more detailed discussion, see Leigh and Nicolai (2013). There are two opposite directions one can take to resolve this entanglement:

(a) to clearly separate the domains (or sorts) of the bearer of truth and the object of the subject matter;
(b) to choose a subject matter whose domain of discourse intrinsically contains both the bearers of truth and mathematical objects altogether.

The first direction is taken by Heck (2011) and Nicolai (2014) for example (theory of truth with disentangled syntax from object-language). Kahle's theories and Eberhard

[18] In contrast to Kahle's FON, Beeson introduces a predicate expressing "x is a proposition" independently as another primitive predicate, and his theory F is based on another type of applicative theory EON. These differences yield no difference in proof-theoretic strength, and Kahle's FON and Beeson's F are in fact proof-theoretically equivalent.

and Strahm's theories, which take applicative theories as their base theories, can be regarded as taking the second direction, since the objects of the subject matter and the bearers of truth coexist in the intended domain of applicative theories.

1.2.2.3 "Unfolding Feasible Arithmetic and Weak Truth" by Sebastian Eberhard and Thomas Strahm

Eberhard and Strahm's contribution extends Feferman's *unfolding programme* to feasible arithmetic, and is a continuation of their previous study of theories of truth for feasible arithmetic (see Eberhard and Strahm 2012; Eberhard 2013).

Axiomatic theories of truth are traditionally based on Peano Arithmetic PA (or its equivalents such as Cantini's OP (1996)), but one may adopt a different kind of base theory. One way to go is to enrich a base theory to, say, a set theory ZF for example; see Fujimoto (2012). Eberhard and Strahm go instead in the opposite direction and weaken the base theory to the so-called feasible arithmetic.

In complexity theory, a branch of theoretical computer science, effective decision procedures or algorithms are classified into a hierarchy of various complexity classes. Some effective algorithms are 'efficient enough' and can be 'feasibly computed', while others are 'too inefficient' and take an intractable amount of time to terminate. Feasibly computable algorithms are often identified with polynomial time algorithms.[19]

Peano Arithmetic PA is too strong a base theory for a theory of truth for feasible arithmetic. This is because the class of definable functions of PA properly includes that of primitive recursive functions, and not all primitive recursive functions are feasibly computable.[20] Over the last few decades, we have seen the development of a variety of formal arithmetic theories associated with the class of feasibly computable functions, in the sense that the class of functions that the theory can 'describe' (in terms of provable totality, provable convergence, definability, etc.) coincides with that of feasibly computable functions. The authors previously presented a theory T_{PT} of truth of feasible strength (Eberhard and Strahm 2012) where the provably total functions are precisely polynomial time computable ones. The present paper presents a new theory $\mathcal{U}_T(\mathsf{FEA})$ of truth (as well as two other proof-theoretically equivalent theories) of feasible strength but this time in the form of Feferman's unfolding.

The notion of *unfolding* was presented by Feferman (1996) as his most recent answer to the following problem raised by Kreisel: 'What principles of proof do we

[19] Roughly speaking, a polynomial time algorithm for a computational problem P is an algorithm such that it can reach the solution of P for any input of length n within $F(n)$ steps of computation for some fixed polynomial function F. Granted this identification of feasibility and polynomial time computability, the $P = NP$ problem, a famous Millennium Prize problem, questions whether or not a certain class of computational problems is feasibly solvable. There are many good textbooks on complexity theory; e.g. see Garey and Johnson (2002).

[20] As a matter of fact, such a base theory must be even weaker than $\mathsf{I}\Sigma_1$, since the class of provably recursive (and thus definable) functions is in this case exactly that of primitive recursive functions.

recognize as valid once we have understood ... certain given concepts?' (Kreisel 1970).[21] Proof-theoretic analysis of the unfoldings of finitist arithmetic and non-finitist arithmetic are already given by Feferman and Strahm in their (2000, 2010) (all of which are significantly stronger than $\mathcal{U}_T(\mathsf{FEA})$). In principle, unfolding is applied to schematic theories which contain schematic axioms expressed in terms of free predicate variables P, Q, ... of each arity; for example, the induction schema is expressed by a single formula:

$$P(0,\vec{v}) \wedge (P(x,\vec{v}) \rightarrow P(x+1,\vec{v})) \rightarrow \forall x\, P(x,\vec{v})$$

by means of a predicate variable P in a schematic theory. The idea behind unfolding is as follows. An initial schematic theory S comes with basic operations and predicates of a subject matter from which we start the process of unfolding. We go on to define and introduce more and more operations and predicates to the initial schematic theory following the rules governed by a certain background theory of operation-forming and predicate-forming. Then, application of schematic axioms is expanded to those newly introduced operations and predicates by means of the Substitution Rule, which allows us to substitute anything (possibly containing new operations and predicates) for the predicate variables P, Q, ... in the schematic axioms of S. In general, unfolding systems comprise: (1) a schematic base theory, which determines the subject matter and universe of discourse of one's investigation; (2) a theory of operation—and predicate-formation, which determines what new operations and predicates we can construct and how they are constructed from the basic ones of S; and (3) a substitution rule, which enables us to apply the schematic axioms of S to newly constructed operations and predicates.

Put this way, it could be said that the essence of unfolding systems lies in the choice of the second component, i.e. its theory of operation—and predicate-formation. Feferman introduced two different types of unfolding: *operational* and *full* unfoldings, which differ in this second component. The former type only allows the introduction of new operations (over individuals); while the latter allows the introduction of both operations and predicates. According to Feferman and Strahm, '[w]hereas [the operational unfolding of S] addresses the question of which operations on $\mathbb{A}$ ought to be accepted given a schematic system S for a structure $\mathcal{A} = (\mathbb{A}, F_0, \ldots, F_n)$, the central question concerning [the full unfolding of S] can be stated as follows: which operations on and to predicates—and which principles concerning them—ought to be accepted if one accepts S?' (Feferman and Strahm 2000, p. 80).

In Feferman and Strahm's formulation in their (2000, 2010), full unfolding systems contain *terms* for 'predicates' for each arity and a binary predicate symbol $\in$ for the membership relation. For an n-ary predicate term X and n-tuple $(i_1, \ldots, i_n)$ of individuals, the formula $(i_1, \ldots, i_n) \in X$ means that $i_1, \ldots, i_n$ fall under the extension of the predicate (expressed by) X. A full unfolding system can thereby treat predicates as terms, and more and more predicates (as terms) are produced by manipulating or

[21] It might be worth noting here that perhaps the most famous axiomatic theory of truth KF was originally presented by Feferman (1991) as an answer to this question of Kreisel's.

combining those terms. Now, one can find here an intimate connection between the treatment of predicates in full unfolding systems and that in truth theories: to say that objects fall under the extension of a predicate is essentially to say that the predicate is true of the objects; cf., Kahle's contribution to this volume. In fact, in his 'pilot study' paper (Feferman 1996), Feferman originally formulated full unfoldings in terms of a truth predicate instead of the membership relation. Following Eberhard and Strahm, let us call this version of full unfolding *truth unfolding*; with respect to finitist and non-finitist arithmetic, truth unfolding and full unfolding are equivalent.[22]

In the present paper, Eberhard and Strahm present the truth unfolding, full unfolding, and operational unfolding of feasible arithmetic, and show that all three systems have the same feasible strength.[23] Through their research, Eberhard and Strahm open up a new subject of study: theories of truth for feasible arithmetic, and they provide a new perspective on the unfolding programme from the point of view of theories of truth.

1.2.2.4 "Some Remarks on the Finite Theory of Revision" by Riccardo Bruni

Paradoxes in general, as Quine noted in his popular (1976), have often stimulated reflection in new directions and have given rise to fruitful new methods and concepts that have been applied to other subjects. Paradoxes of truth are no exception here, and Bruni's essay can be seen as yet another application of the revision rule inspired by the revision theory of truth. The revision theory of truth, invented by both Gupta and Herzberger independently, and developed by Gupta and Belnap (1993), is one of the main contenders in the search for a solution to the Liar paradox. Revision theory identifies the source of the paradox as residing in the fact that the truth predicate admits a circular definition and pathological sentences are to be expected in the presence of circular concepts. Using an example from Gupta and Belnap (1993), suppose we define a predicate Gx as $x = \text{Socrates} \vee (x = \text{Plato} \wedge \neg Gx)$. From a classical perspective, no extension can be assigned to G, since to determine an extension we need to determine which elements of the domain satisfy the *definiens*, and, since G itself occurs in the *definiens*, we already need the extension of G in order to do that. However, the key intuition behind revision theory is that the circular definition gives us *hypothetical* information about the extension of G, and that this

[22] The full unfolding system in the formulation in Feferman and Strahm (2000, 2010) has one more important facility: the disjoin union operator *Join*, for predicate-formation. Without the *Join* operator, the resulting unfolding does not reach the strength of truth unfolding (and full unfolding) with respect to non-finitist arithmetic. In contrast, as Eberhard and Strahm show in the present paper, *Join* yields no difference in proof-theoretic strength with respect to feasible arithmetic (this is also the case with respect to non-finitist arithmetic).

[23] Eberhard and Strahm's theories are based on a certain type of applicative theory, just as Kahle's theories are; see the previous section introducing Kahle's paper for a discussion of the philosophical import of the applicative setting in comparison to the traditional setting. In general, applicative base theories are a natural set-up for pursuing the unfolding programme, because they provide us with a versatile and natural framework for term application and thus for operation- and predicate-forming.

information can be used to provide a rich theory of the content of G. The circular definition tells us at least this: assuming by hypothesis that Plato is not G, then both Socrates and Plato would satisfy the definiens of G and no other object would; however, assuming that Plato is G, then only Socrates would satisfy the definiens of G. Revision theorists keep track of this information as a rule of revision: if, by hypothesis, the extension of G is {Plato}, then it should be revised to {Socrates}; if it is {Aristotle}, then it should be revised to {Socrates, Plato}, etc. We see that Plato behaves, with respect to this definition, as the Liar does with respect to truth: if it is G, it should be not G and, if it is not G, it should be G.

The second key component of revision theory is the recipe for extracting *categorical* information from the revision rule associated with a circular definition. This is achieved by iterating the process of revision and paying attention to the sentences that have a fixed truth value when the process advances, no matter what the initial hypothesis. In our simple example, it is clear that, for any hypothesis, after the first revision, Socrates is classified as G, all other objects except Plato are classified as not G, and Plato is pathological. Hence, we could categorically assert 'Socrates is G' and 'Aristotle is not G', and we should refrain from asserting either 'Plato is G' or 'Plato is not G'. In general, if we have a language L with interpretation M and domain $|M|$, an expanded language L^+ is obtained from L by adding a new predicate G and a definition $Gx =_{\text{def.}} A_G(x, G)$, where $A_G(x, G)$ is a sentence in L^+. We then define the revision rule δ_A as follows: for any hypothesis H (i.e. for any subset of $|M|$), $\delta_A(H) = \{a \in |M| : (M, H) \models A_G(\overline{a}, G)\}$, where $\overline{a}$ is a constant that names a and (M, H) is the interpretation of L^+ obtained from M by adding H as the interpretation of G. Once we have a revision rule, a revision evaluation sequence is defined as $\delta_A^0(H) = H$, $\delta_A^{n+1}(H) = \delta_A(\delta_A^n(H))$. The revision process can be projected into the transfinite ordinals, defining a limit rule. There are several options available in the literature to do this. The key notion in extracting categorical content from a circular definition is the notion of a reflexive hypothesis. A hypothesis H is n-reflexive if $\delta_A^n(H) = H$, and it is reflexive if it is n-reflexive, for some $n > 0$. Once the revision process arrives at a reflexive hypothesis, all the subsequent iterations of the revision rule also produce reflexive hypotheses and form a cycle that repeats itself indefinitely. This is why Gupta and Belnap consider reflexive hypotheses the best candidates to determine the extension of G, and define a sentence B of L^+ as valid in M if it is true in all interpretations (M, H), where H is a reflexive hypothesis. B is valid if it is valid in all interpretations of L. In the general case, the notion of reflexivity has to be extended to a transfinite ordinal. (Gupta and Belnap 1993 consider other notions of validity.)

Bruni's paper analyses the class of finite definitions, a special class of circular definitions that satisfy the condition that for all interpretations M, there is a number k such that, for every hypothesis H, $\delta_A^k(H)$ is reflexive. Finite definitions are those that guarantee arriving at a reflexive hypothesis in a finite number of steps of revision for any hypothesis. The main source of examples of finite definitions is game theory, where the rational action for any player depends on what it is rational for the other players to do.

Bruni's paper evaluates finite definition semantics and compares it with standard (transfinite) revision semantics, highlighting three aspects. Firstly, finite revision semantics has less complexity than standard revision semantics. As an example of this, Bruni proves that every definable set on a circular predicate in the standard interpretation of arithmetic is at most $\prod^1_1$ in finite revision semantics; while (as proved by P. Welch 2003) it is at least Δ^1_2 in the transfinite case. Secondly, finite revision semantics has a sound and complete natural deduction calculus (due to Gupta and Belnap 1993) and Bruni presents an equivalent Hilbert calculus; while no complete calculus can be given for standard revision semantics. Thirdly, these calculi are not only technically interesting, but are also very natural, since they reflect the ordinary arguments one would make when reasoning with circular definitions.

Even though the scope of finite revision theory does not include the truth predicate, in the last part of the paper Bruni develops earlier work by Halbach (1994) that establishes a connection between truth as formalized in FS, and validity as codified in finite revision semantics. Bruni presents a syntactic version of Halbach's connection, showing that (when working in standard arithmetic) derivations from FS can be mimicked in a variation of the theory FS that uses indexed formulae, where the indices represent the stages in the revision evaluation sequence. These results raise the question of whether similar connections can be established in transfinite revision semantics.

1.2.3 Truth as a Substantial Notion

The first two papers in this chapter address common objections against substantial notions of truth and thereby pave the way for inflationary, as opposed to deflationary, theories of truth.

Sher addresses one of the main difficulties that correspondence theories traditionally face, which is none other than the need for a precise construal of the correspondence relation in a way that does not restrict the notion of truth to a single domain of discourse. In order to meet this requirement, Sher proposes what she calls 'composite' as opposed to 'direct' correspondence. She illustrates what this composite relation comes down to in the philosophy of mathematics; a most challenging domain for the correspondence theorist.

Glanzberg adopts an inflationary contextualist approach and argues that hierarchical accounts are necessitated by a reflection process that is meant to render explicit what is involved in our implicit grasp of the notion of truth. Alongside this general motivation, Glanzberg addresses specific arguments that have been put forward against hierarchical approaches to truth. Among these is what he calls the 'one concept' objection, which threatens the unity of the notion of truth; albeit in a very different way from the objection against correspondence theories of truth that Sher addresses in her paper.

Finally, rather than discussing specific arguments against substantial theories of truth, Engel shifts the burden of proof onto the other side by questioning the plausibility of construing the notion of truth as a non-substantial notion. Engel underlines the irreducibly normative role of truth in determining standards of correctness for belief and assertion. He subsequently identifies a tension for the deflationist who will not be able to account for such standards of correctness without admitting the non-deflationary normative content of truth.

1.2.3.1 "Truth as Composite Correspondence" by Gila Sher

A problem for traditional correspondence theories of truth is that it is hard to see how any precise expression of the correspondence relation between discourse and the world could ever account for the property of truth in its multifarious applications, from physical discourse to ethics and mathematics. Arguably, if a direct correspondence account is available for the truth of discourse about the physical world—with terms referring to physically identified objects and the truth of sentences built on reference as Tarski taught us (see also Field 1972)—such an account will give rise to insuperable difficulties when it comes to domains such as ethics and mathematics. It is not surprising then that, especially in mathematics, many philosophers have been tempted to give up on the idea that the truth of discourse amounts to correspondence with mathematical facts. Correspondence theorists thus face a problem. If they stick to the thesis that truth is correspondence, they will have a hard time providing a plausible account of correspondence that applies uniformly to the various realms of discourse. If, however, they admit that truth is correspondence in some domain, and not in others, then they compromise the unity of the notion of truth. So, either there need to be different meanings of the word "true", or "true" cannot be applied to the various domains of discourse to which we ordinarily, and unproblematically—or so it seems—apply it.

In her "Truth as Composite Correspondence", Gila Sher takes up the correspondence challenge in an attempt to overcome the above predicament. She does so in two ways. Firstly, by articulating the main lines of a renewed methodological programme for the development of a substantive theory of truth. The challenge here is twofold: (a) to alleviate what the author argues is the unjustified burden placed by a foundationalist stance on the possibility of developing any substantive account of truth; (b) to make room for an alternative construal of correspondence through a long-term holistic inquiry that would be faithful to the specifics of the various ways we access the world in different domains of discourse. Secondly, Sher sets her programme to work in the case of mathematical discourse. The author's main thesis is that one can conceive of mathematical discourse as being about facts of the world. But what kind of facts? Building on her earlier work[24], Sher argues that these facts consist of the world having some formal properties, where formal properties are in

[24] See Sher (1991).

turn explained in terms of invariance under some classes of transformations. In a nutshell, the invariance idea is here a generalization (a cross-domain generalization) of the idea put forward by Tarski (1986): just as the property of being red, say, can be seen as the property which is invariant under those transformations of the domain of the universe that leave red things red, so the formal properties, such as the (second-order) property of having cardinality 3, are properties that are invariant under *any* permutation of the domain. That much having been said, it remains problematic that mathematical discourse is, on the face of it, about individuals, namely numbers, and not about second-order properties. This is where composite correspondence comes in. First-order statements about numbers can, Sher argues, be said to correspond to facts involving second-order properties, via e.g. posits, if one allows for composite correspondence. And in fact, there are reasons for humans to have adopted a standard of composite correspondence rather than direct correspondence as a substantive norm for their discourse; namely, simplicity or cognitive tractability. There is thus some evidence that a standard of truth understood as composite correspondence is a plausible one. In the remainder of her paper, Sher goes on to sketch reasons for the fruitfulness of her approach in solving various puzzles in the philosophy of mathematics.

1.2.3.2 "Complexity and Hierarchy in Truth Predicates" by Michael Glanzberg

After Kripke's famous attack on the Tarskian hierarchy of languages, hierarchies have been viewed with suspicion. Kripke (1975) argued that a Tarskian hierarchy of truth predicates cannot be a good formalization of ordinary language, because it leaves out perfectly natural self-referential sentences. Even in theories with an untyped truth predicate—that is, theories especially designed to overcome these expressive problems (such as Kripke's own theory)—the eventual reappearance of hierarchies due to the revenge paradoxes is usually taken to be a defect that future research should overcome. In his "Complexity and Hierarchy in Truth Predicates" Glanzberg tries to dispel these worries by offering a sustained defence of hierarchies, and showing where and why they should be expected to appear.

The paper argues that inflationary theories of truth motivate the use of hierarchies, while deflationist theories do not. The key element that generates this difference is that while for deflationists truth is a simple property that is fully characterized by the transparency of truth (the intersubstitutability of any sentence A and $T(\ulcorner A \urcorner)$ in all non-opaque contexts), for inflationary theories truth is a potentially complex semantic property with internal structure. The complexity of truth originates in the different mechanisms involved in the determination of the truth value of sentences: semantic composition and facts about reference and satisfaction. Even if truth is a complex property, ordinary speakers have an implicit grasp of the truth predicate and Glanzberg argues that philosophers can make this implicit knowledge explicit by a process of reflection on our own abilities. The main claim of the paper is that this activity of reflection generates hierarchies.

The central part of the paper provides an overview of several processes of reflection, showing that all of them reveal the complexity of truth and some hierarchy is generated. The first case starts with the language of arithmetic, which one can reflect upon either model-theoretically (if one understands the language as interpreted in a model of arithmetic and then gives a Tarskian definition of truth) or proof-theoretically (if one starts with a theory such as PA and gives the compositional axiomatic truth theory CT). Complexity measures show that the complexity of the truth predicate obtained using a Tarskian definition of truth is greater than the complexity of the base theory. The second case is a language that contains its own truth predicate. From a model-theoretic perspective, Glanzberg here summarizes Kripke's construction of a fixed-point semantics as a process of approximation (what he calls the long iteration strategy for reflection), which shows up in almost all theories designed to provide a solution to the Liar paradox; such as revision, paraconsistent or paracomplete theories. The long iteration strategy is not able to produce a perfect theory of truth. Taking Kripke's construction as an example, when we get to a fixed-point interpretation for the truth predicate we see that the Liar sentence is neither in the extension nor the antiextension of the truth predicate. But then the Liar sentence is not true and we are back to a paradox. This means that the process of reflection is incomplete and has to start again, creating an open-ended hierarchy of new truth predicates. Glanzberg briefly presents his own contextualist solution to the Liar paradox as an example of this process of the generation of hierarchies and points to some recent results on the iteration of axiomatic theories of truth through suitable proof-theoretic ordinals. The fact that a completely successful theory of truth is not found is to be expected, given the extreme complexity of truth.

The last part of the paper defends the hierarchical approach from some common objections: the *one concept* objection (we only have one concept of truth, not many concepts generated at the different levels in the hierarchy), the *clumsiness* objection (some hierarchical theories cannot express some ordinary self-referential sentences) and the *weakness* objection (hierarchical theories of truth are very weak when it comes to mathematical purposes). The discussion of the one concept objection is of special interest. Here Glanzberg introduces the notion of stratification: "A concept is stratified if we cannot provide a single theory or definition for it. Instead, we provide a family of related theories or definitions, each of which is systematically connected to others. In effect, a concept is stratified if when we try to analyze it, we wind up with a hierarchy" (p. 211). Glanzberg compares the case of truth to other notions which are also stratified (such as the notion of mathematical proof), and finds that the one concept objection applies differently to different types of truth hierarchies. Finally, Glanzberg discusses how his own hierarchy stands with respect to the one concept objection. The general conclusion of the paper is that inflationary accounts of truth motivate the construction of hierarchical theories of truth; theories that are more natural than non-hierarchical ones.

1.2.3.3 "Can Deflationism Account for the Norm of Truth?" by Pascal Engel

Many influential deflationists, such as Quine and Field[25] have endorsed their conception of truth as part of a more general physicalist, or naturalist, philosophical framework. What these authors are primarily interested in is whether there is a place for a notion of truth in the language of natural science that would be suitable for the description or causal explanation of natural phenomena. Normative facts are not part of the picture—they are simply not natural phenomena. Science identifies a phenomenon in natural terms and seeks causal explanations; it is indifferent to the explanation of non-natural facts and of behaviour in terms of underlying reasons and norms. Against this backdrop, deflationists, with notable exceptions[26] however, have not paid much attention to the normative role of truth and the problem this role may pose to a general deflationary approach.[27] For, in holding that all there is to our understanding of the notion of truth is our understanding of the assertability of all of the T-sentences, the deflationist implies, among other things, that truth is not a distinctively normative property. It may be true that the notion of truth is not essentially a normative notion. By ascribing truth to the proposition that snow is white we do not thereby make a normative claim—as we would when ascribing, for example, goodness—but the notion of truth is still deeply involved, or so it seems, in explaining the norms that govern some of our actions and thoughts.

In particular, it seems to be part of our understanding of the norms that govern correct assertion and belief that they are correct only if true. Consequently, the notion of truth, even if not in itself normative, is in fact needed in order to account for the norm of correctness governing belief and assertion. This would perhaps not do much harm to the deflationist if it could be shown that our recourse to truth in this case is not to do with setting a distinctive and irreducible norm, and that truth is involved only in a shallow way; for instance, as a mere logical device. In his essay, Engel shows how uncomfortable the deflationist's position is here. As Engel recalls, there are many arguments showing that the standard of truth for assertion (and belief) is distinct from other standards, such as subjective standards (it is correct to assert that p if one believes that p) or epistemological standards, such as warranted assertability. At the same time, difficulties are lurking nearby if one chooses to renounce such a correctness norm. Engel sets out to evaluate the various deflationist strategies which downplay the normative role of the notion of truth and have been proposed in response to these challenges. His conclusion is that the following dilemma is robust: either deflationists have to eliminate the normative features of truth, but then they are unable to account for what constitutes the correctness of belief and assertion; or they grant that truth is involved in the normative account of belief and assertion, but then they are unable to account for the distinctive substantial norms intrinsically associated with truth.

---

[25] See e.g. Quine (1970, 1990) and Field (2001).
[26] See e.g. Horwich (2006). On a larger scale, see also Robert Brandom's work (Brandom 1994).
[27] One such problem was already pointed out early on by Dummett (Dummett 1959).

1.2.4 Deflationism and Conservativity

Halbach and Horsten are concerned with the general norms that a theory of truth should adhere to. Specifying such norms obviously depends on the notion of truth that one endorses, and for Halbach and Horsten those norms are motivated by their deflationary approach to truth. While Engel (see previous section) is interested in the way truth relates to other notions—namely, assertion and belief—Halbach and Horsten assume a perspective that is internal to a particular notion of truth, that is, reflexive truth, and propose a short list of general desiderata that axiomatic theories aimed at accounting for reflexive truth must satisfy. Their viewpoint is normative but also descriptive in that they are interested in describing what drives current work on truth theories. They do not propose their list in order to single out one theory as the best theory currently available; but they do hope that the norms will make it possible to compare theories and shed light on the choices made.

Deflationism famously advocates the non-substantial character of truth. One way to explicate the 'non-substantiality' thesis has been to take it to mean that a theory of deflationary truth should be conservative over its base theory, which in this context is commonly taken to be Peano Arithmetic (PA). A theory of truth is conservative over its base theory if it proves no theorems in the language of the base theory which are not already provable in the base theory alone; otherwise, it can be argued that the truth theory adds substantial content to that of the base theory. Conservativity as a criterion for deflationism was explicitly proposed by Horsten (1995), Ketland (1999) and Shapiro (1998) and is reminiscent of a long history of uses of this notion in the foundations of mathematics.[28]

Leigh's paper 'Some Weak Theories of Truth' follows this line of thought and examines which out of twelve principles of truth considered by Friedman and Sheard (1987) is to be blamed for non-conservativity over (PA). The truth theories that Friedman and Sheard consider are maximally consistent sets of these twelve principles modulo a theory $\mathsf{Base}_{\mathrm{T}}$. All of these theories, except one, are non-conservative extensions of (PA). Leigh observes that one of the principles, namely, U-Inf (the predicate version of the Barcan formula), is present in all of them and that in its absence, conservativity is restored. It follows that this is the principle responsible for the non-deflationary syntactic behaviour of the truth theories.

Fischer's paper 'Deflationism and Instrumentalism' changes the rules of the game somewhat by proposing a new understanding of deflationism and with it, a new way of assessing the conservativity requirement. Fischer construes deflationism as a form of instrumentalism in Hilbert's spirit[29]: carrying out the programme of instrumental deflationism means showing that it is possible to combine conservativity (i.e. truth-theoretic innocence) with instrumental utility (given by the expressive power of speed-up results). Fischer uses a weak theory of truth, PT^{-}, to show that this is possible. On top of this novel endorsement of the conservativity requirement,

[28] See e.g. Hilbert (1926), Field (1980) and Shapiro (1983).
[29] This understanding also underlies Ketland (1999).

Fischer offers additional support by addressing one of the main objections against conservativity according to which the truth theory should prove the soundness of its base theory, here PA.

Both Leigh's and Fischer's papers consider conservativity in its proof-theoretic sense, that is, they are concerned with the deductive power of the truth theory. In his 'Typed and Untyped Disquotational Truth', however, Cieśliński draws our attention to *semantic* conservativity and its relation to proof-theoretic conservativity for disquotational theories of truth, which are deflationary theories par excellence. It has been argued that semantic conservativity fits the deflationary spirit better (McGee 2006) as it evidences the lack of extra metaphysical commitments and for this reason it can be seen as an expression of the demand for an innocent non-substantial notion of truth. Cieśliński shows that conservativity in this sense is difficult to obtain for disquotational theories of truth. In preparation for this argument, Cieśliński offers an overview of disquotationalism and the problems inherent in devising a disquotational theory of truth, which have to do with deciding on adequate criteria for selecting a consistent subset of T-biconditionals, given that the full T-schema cannot be upheld.

Reading Enayat and Visser's paper 'New Construction of Satisfaction Classes' reminds us that the question of conservativity is also pressing for typed, and not just for untyped, theories of truth. The paper takes us back to Tarski's definition of truth in a model, for which the notion of satisfaction was famously introduced. As an axiomatization of this, Enayat and Visser propose $\mathsf{PA}^{\mathsf{FS}}$ as the base theory: a theory with a satisfaction predicate added to PA. It is reasonable to expect an axiomatic theory of Tarskian satisfaction to be conservative over its base theory and indeed the deflationist would be in an awkward position if conservativity fails in this case (although note that Enayat and Visser themselves do not draw any connections between their work and deflationism). However, the model-theoretic proof of conservativity (via the completeness theorem) with PA as the base theory is not easy, roughly because not every model of PA can be expanded to a model of $\mathsf{PA}^{\mathsf{FS}}$. Enayat and Visser offer a new, more simplified proof which is a clear improvement on previous results in this area.

1.2.4.1 "Norms for Theories of Reflexive Truth" by Volker Halbach and Leon Horsten

Given the proliferation of axiomatic truth theories in the last two decades, the question naturally arises of how one should adjudicate between them. Earlier attempts to answer this question include Sheard (1994) and Leitgeb (2005). The present paper extends that line of research by proposing a list of norms for axiomatic theories of the reflexive use of truth (or 'type-free' truth), as opposed to other uses, e.g. the use of truth in natural language. It is, of course, because there is no obvious way to circumvent the Liar paradox that there are many axiomatic truth theories suggested in the literature.[30]

[30] See also Halbach (2010) for an exposition of axiomatic truth theories.

The authors first offer some methodological remarks concerning the project they undertake. They doubt that a single property or cause can be identified from which all truth norms can be derived. However, they do not see the various desiderata as independent of each other, as they believe that satisfying truth-norms is not an all-or-nothing affair and that norms can be satisfied to a lesser or greater degree by different theories. This is where the authors distance themselves from the approach of Leitgeb (while mostly agreeing on the norms themselves) who considers truth theories as maximally consistent subsets of the norms that he lists.

Halbach and Horsten intend their norms to be underdetermined but also their list to be exhaustive in the sense that any desirable feature of an axiomatic theory of truth is somehow derived from the list. For example, in comparing their list to Sheard's, the authors take the necessitation rule to be a special case of the disquotational requirement, and the inference rule $\phi \rightarrow \psi \vdash T^{\ulcorner}\phi^{\urcorner} \rightarrow T^{\ulcorner}\psi^{\urcorner}$ as being derived from the compositionality norm and the identity between inner and outer logic, which also follows from the disquotational requirement. Since their list is very short, most of the work goes into determining which features of theories of truth are derived from which desiderata. An advantage of this approach is that it is open to novel ways (properties of the truth theory) of satisfying the desiderata. However, choosing what is fundamental and what is derivative is directly influenced by one's philosophical views. Under the deflationist approach that the authors adopt, the identity between inner and outer logic follows from the disquotational requirement and need not be listed as a separate norm; as it is, for example, by Leitgeb.

These, briefly are the norms given by Halbach and Horsten:

1. *Coherence*. A truth theory may be incoherent in relation to its base theory, if, for example, it contradicts theorems of the latter, or it is ω-inconsistent, or the induction schema cannot be extended to the language with the truth predicate. The truth theory may also be incoherent in its truth-theoretic part, if, for example, it proves $T^{\ulcorner}\phi^{\urcorner}$ for all ϕ.
2. *Disquotation and Ascent*. This norm stems from the deflationist's conception of truth as a disquotational devise or a devise for performing semantic ascent, and it entails that sentences ϕ and $T^{\ulcorner}\phi^{\urcorner}$ are in some sense equivalent. Ways to make this equivalence precise are (i) through the idea of transparency (Field 2008); meaning that ϕ and $T\phi$ are intersubstitutable without cost, or (ii) using the T-biconditionals; that is, material equivalences of the form $T(^{\ulcorner}\phi^{\urcorner}) \leftrightarrow \phi$. Both these ways lead to paradox under very weak assumptions, which means that the disquotational requirement cannot be met in full. The question then is how to weaken the disquotational requirement while still obtaining a non-trivial theory of truth. To restrict the T-schema, a straightforward proposal is to admit as many of its instances as possible. However, McGee (1992) shows that one cannot single out a unique maximally *consistent* set, and there is currently no uniform principle available to select a unique maximal set of admissible instances of the T-schema. To restrict the T-biconditionals, an obvious choice is to offer an alternative for the material conditional $\leftrightarrow$. Within classical logic the T-biconditionals may be consistently replaced by inference rules; that is, the truth theory may be closed under Necessitation $S \vdash \phi \Rightarrow S \vdash T(^{\ulcorner}\phi^{\urcorner})$ and Co-necessitation $S \vdash T(^{\ulcorner}\phi^{\urcorner}) \Rightarrow S \vdash \phi$. The possibilities proliferate in the case of

non-classical logics.
3. *Compositionality*. This norm requires that the T-predicate commutes with the connectives and the quantifiers (at least for vagueness-free fragments of the language). In itself commutation is not sufficient; one also needs the T-biconditionals for at least the atomic sentences in the ground language in order to get compositionality for the T-free fragment. Yet, it may not be possible to achieve full compositionality as it may clash with other norms. Motivation for restricting compositionality may then come from taking truth to be a partial concept; for example, one that does not apply to meaningless sentences such as the Liar paradox. One then naturally wants to reject commutation of the T-predicate with negation. Otherwise, negating the truth of the Liar sentence would mean that the negation of it is true, while this negation should be as meaningless as the Liar sentence itself. This is how things go with theories such as KF or Burgess' theory (Burgess 2009), where the restriction to positive compositionality gives rise to a grounded notion of truth. So groundedness is in some cases a derivative from the more general requirement of compositionality.
4. *Sustaining ordinary reasoning*. Feferman (1984) famously stated that a truth theory should sustain ordinary reasoning. There are many ways to interpret this norm; it rules out, for example, logics that do not allow the truth predicate in the induction scheme.
5. *A philosophical account*. This is a meta-norm that demands that the proposed norms are philosophically justified; since no truth theory can satisfy all norms in full, a philosophical story is needed to justify why certain axioms have been chosen and not others. Kripke's construction of the extension of the truth predicate as a learning process can be seen as one such philosophical story which justifies grounded truth. The story may not apply to other theories, and philosophical justification should be compatible with 'truth-theoretic pluralism' for the uses of the truth predicate in different contexts. The idea of truth-theoretic pluralism for axiomatic theories is already found in Sheard (1994) under the name of 'local truth analysis'. Finally, Halbach and Horsten add that, although desirable, this last norm should not hinder research on truth theories that is not motivated by the aim of providing a philosophical account for a particular use of the truth predicate.

1.2.4.2 "Some Weak Theories of Truth" by Graham E. Leigh

The question of whether a theory of truth is conservative or not over a base theory has acquired philosophical significance in the discussion of the nature of truth and the debate over deflationary conceptions of truth. In this connection, Leigh's essay on *Some Weak Theories of Truth* helps to circumscribe the theoretical possibilities that are open to the conservative deflationist, as it determines the nine *maximally conservative* sets of a collection of twelve principles of truth modulo a fixed theory $\mathsf{Base}_{\mathrm{T}}$. His research is motivated by the question: 'What assumptions about the nature of truth are responsible for deciding the proof-theoretic strength of a theory of truth?'. This motivation leads him to extend Friedman and Sheard's programme (Friedman and Sheard 1987).

Friedman and Sheard (1987) list twelve principles of truth, each of which is quite natural and plausible, but which taken together are inconsistent, and determine the nine *maximally consistent* combinations of them over a fixed theory $\mathsf{Base}_{\mathrm{T}}$ of truth including PA. The theory $\mathsf{Base}_{\mathrm{T}}$ is formulated over the language $\mathcal{L}_{\mathrm{T}}$ (the language of arithmetic plus a truth predicate T), and its axioms consist of those of PA with the expanded arithmetical induction schema for $\mathcal{L}_{\mathrm{T}}$ and the following three truth-theoretic axioms[31]:

(i) $\quad \forall\phi\forall\psi[T(\phi \to \psi) \to (T\phi \to T\psi)]$;

(ii) $\quad \forall\phi[(\phi \text{ is an axiom of } \mathsf{PRA}) \to T\phi]$;

(iii) $\quad \forall\phi[(\phi \text{ is a logically valid } \mathcal{L}_{\mathrm{T}}\text{-formula}) \to (\text{the universal closure of } \phi \text{ is true})]$.

Besides these basic axioms, Friedman and Sheard list twelve further truth-theoretic principles; see Table 1 in Leigh's paper. Then, they specify the following nine maximally consistent combinations $\mathcal{A}$–$\mathcal{I}$ of the twelve principles modulo $\mathsf{Base}_{\mathrm{T}}$:

$\mathcal{A}$. In, Intro, ¬Elim, Del, Rep, Comp, E-Inf, U-Inf.
$\mathcal{B}$. Rep, Comp, Cons, E-Inf, U-Inf.
$\mathcal{C}$. Del, Comp, Cons, E-Inf, U-Inf.
$\mathcal{D}$. Intro, Elim, ¬Intro, ¬Elim, Comp, Cons, E-Inf, U-Inf.
$\mathcal{E}$. Intro, Elim, ¬Intro, Del, Cons, U-Inf.
$\mathcal{F}$. Intro, Elim, ¬Elim, Del, U-Inf.
$\mathcal{G}$. Intro, Elim, ¬Elim, Rep, U-Inf.
$\mathcal{H}$. Out, Elim, ¬Intro, Del, Rep, Cons, U-Inf.
$\mathcal{I}$. Elim, ¬Elim, Del, U-Inf.

For each $\mathcal{X}$ among $\mathcal{A}$–$\mathcal{I}$, $\mathcal{X} + \mathsf{Base}_{\mathrm{T}}$ is consistent but adding any more of the twelve principles to $\mathcal{X} + \mathsf{Base}_{\mathrm{T}}$ results in an inconsistent theory, and every consistent (over $\mathsf{Base}_{\mathrm{T}}$) set of the 12 principles is included in some of $\mathcal{A}$–$\mathcal{I}$. In what follows, we identify each $\mathcal{X}$ and its induced theory $\mathcal{X} + \mathsf{Base}_{\mathrm{T}}$ for simplicity. The proof-theoretic analysis of these systems is given by Cantini (1990), Halbach (1994), and Leigh and Rathjen (2010), and their proof-theoretic ordinals are all known. Leigh and Rathjen also studied the Friedman-Sheard programme in intuitionistic logic; see Leigh and Rathjen (2012) and Leigh (2013).

Conservative theories of truth have a special status in deflationism. Among Friedman and Sheard's nine maximally consistent theories, only $\mathcal{A}$ is conservative over PA, and all the others go beyond PA; hence, if one accepts the conservativity requirement, $\mathcal{B}$–$\mathcal{I}$ are not deflationist theories of truth. So, which principle is responsible for the non-deflationary deductive power of these systems?

First, it is noted that the nine maximally consistent theories $\mathcal{A}$–$\mathcal{I}$ all contain the principle U-Inf. Also, E-Inf always comes together with Comp. Now, we can easily

[31] It is debatable whether $\mathsf{Base}_{\mathrm{T}}$ is a necessary basic part of theories of truth or even whether it is acceptable or not; some influential and popular axiomatic theories of truth such as KF (Feferman 1991) and DT (Feferman 2008) are inconsistent with the three axioms (i)–(iii) of $\mathsf{Base}_{\mathrm{T}}$.

verify that U-Inf implies E-Inf over $\mathsf{Base}_{\mathrm{T}}+\mathsf{Comp}$. Consequently, the following list gives the combinations of the principles that induce maximally consistent theories over $\mathsf{Base}_{\mathrm{T}}$ + U-Inf (rather than over $\mathsf{Base}_{\mathrm{T}}$):

A. In, Intro, ¬Elim, Del, Rep, Comp.
B. Rep, Comp, Cons.
C. Del, Comp, Cons.
D. Intro, Elim, ¬Intro, ¬Elim, Cons, Comp.
E. Intro, Elim, ¬Intro, Del, Cons.
F. Intro, Elim, ¬Elim, Del.
G. Intro, Elim, ¬Elim, Rep.
H. Out, Elim, ¬Intro, Del, Rep, Cons.
I. Elim, ¬Elim, Del.

Leigh's paper demonstrates that B–I yield a conservative theory over PA when they are adjoined to $\mathsf{Base}_{\mathrm{T}}$ (rather than $\mathsf{Base}_{\mathrm{T}}$ + U-Inf). This result indicates that U-Inf is responsible for the non-deflationary deductive power of $\mathcal{B}$–$\mathcal{I}$: that is to say, the culprit is spotted!

Leigh's paper sheds light on the subtle interactions between principles of truth and their effects on the truth-free consequences of a theory. For example, his results naturally raise the following interesting question: why does U-Inf add proof-theoretic strength? One may find a certain contrast between 'compositional' and disquotational principles here; U-Inf and its dual E-Inf may be called 'compositional' principles of truth because they partially axiomatize the compositional nature of truth where the truth of a sentence depends on the truth (or semantic value) of the constituents of that sentence; whereas the other principles, except Cons and Comp, are disquotational in the sense that they capture a fragment of the T-schema (for $\mathcal{L}_{\mathrm{T}}$).

1.2.4.3 "Deflationism and Instrumentalism" by Martin Fischer

At the root of deflationary conceptions is the suggestive, yet vague, idea that the notion of truth is 'useful' but not 'substantial'. This understanding of deflationism is underspecified as it leaves room for a variety of explanations, and indeed, deflationists disagree as to how to make it precise. On one understanding, truth is an 'expressive device' but not a natural property; on another, it is not a property at all. Some say that truth has no causal-explanatory force, while still others claim that it has no explanatory force whatsoever. Some think that all that is essential to our understanding of the concept of truth is our acceptance of Tarski's T-schema or some subset of its instances; others that the notion of truth is essentially a logico-syntactic device. Most of these claims, however, proved to be sufficiently vague to make it hard to assess, prove or refute them conclusively. To remedy this situation logical tools have had an increasing presence in discussions of deflationism in recent years. Philosopher-logicians, following the lead of Tarski, have discussed what would count as precise 'adequacy conditions' for a formalized theory, such that they would serve

as explications, in Carnap's sense, of the deflationists' claim. Such adequacy conditions are typically meant to ensure that the theory is a theory of *truth* (this is also a way to understand Tarski's material adequacy condition), that the theory accounts for a class of expected uses of the truth predicate (which account for its 'usefulness') and that truth is 'non-substantial' or innocent. The conclusions have not been overly favourable to the deflationist in that on the one hand, it has been claimed, against deflationism, that on some plausible precise explication of 'usefulness' and 'non-substantial' deflationary theories of truth cannot exist; while on the other hand, the deflationists have been slow to provide alternative consistent adequacy conditions.

In his essay, *Deflationism and Instrumentalism*, Fischer takes up the challenge on behalf of the deflationist. His philosophical starting point is the idea that deflationism about truth should be construed as a branch of instrumentalism. According to this view, vindication of deflationism depends on the possibility of carrying out an instrumentalist programme conceived in the spirit of Hilbert's programme that aims to show both the instrumental utility and the theoretical innocence of the truth predicate. In this paper, Fischer argues that this programme can be carried out successfully. The core of the argument is the proof that it is possible to devise a truth theory (Fischer's favourite being PT^-) which is conservative over PA (conservativity for innocence), and at the same time allows significant epistemic benefits by shortening of proofs (speed-up for epistemic usefulness). By proving conservativity and speed-up results for PT^-, Fischer lends credibility to his instrumental deflationism. Perhaps this would not have been entirely satisfactory if Fischer had not, in the same paper, also addressed the now classical criticism of conservative theories of truth due to Shapiro (1998) and Ketland (1999). The criticism is based, roughly speaking, on the notion that theories of truth should prove the soundness of their base theories and that a conservative theory of truth cannot do so (by Gödel's second incompleteness theorem). In the last section of his paper, Fischer argues that this criticism is not conclusive. Fischer does not directly take issue with the claim that an adequate theory of truth should prove soundness. Rather, he insists that, as is well known to logicians, the impossibility result strongly relies on soundness being formulated in the form of strong reflection principles. There are other ways to express soundness, however, and Fischer argues that some of them are such that: firstly, they constitute acceptable formulations of soundness; and secondly, a truth-theoretic conservative extension of PA can prove the soundness of PA so expressed. If Fischer is right in his conclusion, then his new way to deal with the 'conservativity argument' deserves serious consideration from both the deflationist and the non-deflationist. More generally, his essay proposes new logico-philosophical standards for assessing the value of various formalized theories of truth from a deflationist perspective, at the intersection between formal and philosophical reflections on truth.

1.2.4.4 "Typed and Untyped Disquotational Truth" by Cezary Cieśliński

Cieśliński's contribution studies a certain type of axiomatic theories of truth called *disquotational theories* of truth. In particular, it focuses on those disquotational

theories that are conservative over their base theory (PA in this setting), and for this reason generally thought to be consonant with the deflationists' claims about the nature and function of truth. This paper shows that there is more to the notion of conservativity that the deflationist ought to consider.

For the sake of simplicity, let us assume that the bearers of truth are sentences, though the following discussion would still apply for standpoints that take other objects—such as proposition—as the bearers of truth.

The core doctrine of *disquotationalism* holds that the content of the notion of truth is thoroughly captured by a certain collection of the so-called T*-biconditionals*, which are the sentences of the following form:

$$\sigma \text{ is true iff } \sigma. \tag{1.1}$$

Let us call statement 1.1 the T-biconditional *for* σ.

Tarski's famous Convention T contends that a predicate T is a materially adequate truth predicate of a language $\mathcal{L}$, when T validates the *full* T*-schema*: i.e.

$$T\ulcorner\sigma\urcorner \text{ iff} \sigma, \text{for } \textit{all} \text{ sentences } \sigma \text{ of the language } \mathcal{L}, \tag{1.2}$$

where $\ulcorner\sigma\urcorner$ denotes the name (or *structural descriptive* in Tarski's terminology) of sentence σ. However, Tarski's undefinability theorem tells us that no language $\mathcal{L}$ can contain a materially adequate truth predicate T for the language $\mathcal{L}$ itself. For, by taking a self-referential sentence λ such that $\neg T\ulcorner\lambda\urcorner$ iff λ, we can immediately derive a contradiction from the particular T-biconditional for λ, which is an instance of the full T-schema.

This fact forces disquotationalists to place restrictions on the full T-schema and impels them to find a suitable proper subschema of the full T-schema by excluding the contradictory T-biconditionals. Cieśliński's paper begins with a brief exposition of the philosophical background of the disquotational approach, and the reader may also refer to Halbach (2010) and Horsten (2011).

Now, as we have seen, at least the T-biconditional for the above particular λ ought to be excluded from any disquotational theory of truth. However, we cannot block inconsistency simply by expelling λ, since many other T-biconditionals yield inconsistency in various manners.[32] Thus the question arises of which subschema of the full T-schema correctly characterize the disquotationalist conception of truth.

An obvious candidate for a suitable disquotational theory of truth is that obtained by restricting the full T-schema to sentences that do not contain any occurrence of T; the resulting theory is often denoted by $\mathsf{TB}\restriction$ (or TB^{-}) in the literature[33], which stands

[32] For instance, let us consider a pair of sentences ρ_0 and ρ_1 such that $T(\ulcorner\rho_0 \leftrightarrow \rho_1\urcorner)$ iff ρ_0 and that $T(\ulcorner\rho_1 \leftrightarrow \neg\rho_0\urcorner)$ iff ρ_1; for a formal construction of such sentences in arithmetic, we refer readers to Boolos (1993) or Hájek and Pudlak (1993). Then, we can easily derive a contradiction from the T-biconditionals for ρ_0 and ρ_1.

[33] When the base theory is PA, the result of furthermore adding the expanded arithmetical induction schema for the expanded language (obtained by adjoining the truth predicate T to the language of PA as a new primitive predicate symbol) to $\mathsf{TB}\restriction$ is more simply called TB.

for "Tarski Biconditionals".[34] The theory $\mathsf{TB}\restriction$ is known to be conservative over its base theory, the proof of which is originally due to Tarski (1983): any consequence of $\mathsf{TB}\restriction$ in the base language can be derived in the base theory without using any T-biconditionals.

This restriction of the full T-schema results in a *typed* conception of truth and thus would be renounced by disquotationalists who like to encapsulate (part of) the self-applicative character of the notion of truth in their theories. Hence, disquotationalists in favour of a self-applicative conception of truth would have to search for another suitable "untyped" subschema of the full T-schema.

It is sometimes wrongly thought that disquotational theories must be deductively fairly weak. This misconception may perhaps have been wrongly deduced from Tarski's conservativity result for $\mathsf{TB}\restriction$ or derived from the redundancy theoretic point of view; redundancy theorists claim that "is true" can be eliminated by means of the equivalence between $T\ulcorner\sigma\urcorner$ and σ and thus "is true" is redundant. However, as is also mentioned in Cieśliński's paper, it follows from McGee's trick (McGee 1992) that any sentence, regardless of whether it contains the truth predicate or not, is equivalent to some T-biconditional over Peano arithmetic, and thus any set of axioms can be reaxiomatized over PA as a disquotationalist theory of truth, i.e. a set of T-biconditionals.[35] Consequently, "untyped" disquotational theories can have arbitrary strength. To make matters worse, McGee (1992) also showed that there are uncountably many mutually inconsistent maximally consistent subschemata of the full T-schema. Furthermore, it follows from Theorem 2 of McGee (1992) that there are uncountably many mutually inconsistent maximally sound (or ω-consistent) subschemata of the full T-schema. Later, Cieśliński (2007) even showed that there are uncountably many incompatible maximally conservative subschemata of the full T-schema. So, how should we select a consistent subschema and what kind of subschema is to be chosen from this vast variety of options? Unless we stick to a Tarskian-type distinction like $\mathsf{TB}\restriction$, the main challenge for the disquotationalist is the problem of specifying a sensible set of T-biconditionals.

Although in principle they are independent of each other, disquotationalism is often correlated to (or even subsumed in) deflationism in the philosophical literature; we may perhaps say that the disquotationalist tenet that the statement "it is true that . . . " is a mere paraphrase of the statement ". . . " gave the prototype of deflationism of truth. For instance, Halbach (2010, § 21) counted disquotationalism as one of the

[34] Alternatively, we may allow *parameters* in T-biconditionals and obtain a disquotational theory of truth by restricting the full T-schema with parameters to the sentences containing no occurrence of T; the resulting theory is often called $\mathsf{UTB}\restriction$ (or UTB^{-}), which stands for "Uniform Tarski-Biconditionals"; for more details, see Halbach (2010). As in the case of $\mathsf{TB}\restriction$, the result of adding the expanded arithmetical induction schema to $\mathsf{UTB}\restriction$ is called UTB, which corresponds to UTB_1 in Cieśliński's paper.

[35] Indeed, it follows from the arithmetized completeness theorem that any recursive consistent theory over any recursive language can be interpreted in some disquotational theory. Hence, for example, disquotational theories (over PA) can have even greater consistency strength than ZF plus the existence of very large cardinals.

core doctrines of deflationism. According to Halbach, another core doctrine of deflationism consists of the insubstantiability of truth: "truth is a thin notion in the sense that it does not contribute anything to our knowledge of the world" (Halbach 2010, p. 310). As mentioned earlier (c.f. § 2.4), some argue that this second doctrine translates into the conservativity requirement, according to which a deflationary theory of truth must yield no consequence that is not already a consequence of the base theory.

A common formal formulation of the conservativity requirement is:

If a theory S of truth over a base theory B is deflationary,

then $\mathsf{S} \vdash \phi$ implies $\mathsf{B} \vdash \phi$ for all sentence ϕ of the language $\mathcal{L}_0$ of B. (1.3)

However, there is another formulation of "conservativity". We say a theory S of truth is *conservative in the semantic sense* (or *semantically conservative* in Cieśliński's paper) over a base theory B iff every model of B is expandable to a model of S: So from this, due to the completeness theorem of first-order logic, conservativity in the semantic sense implies conservativity in the ordinary sense (or in the proof-theoretic sense (McGee 2006)); while the converse does not generally hold. For example, McGee (2006) argues that conservativity in the semantic sense is preferable from the deflationist point of view and has more philosophical significance, since a move from a base theory B to a conservative theory of truth over B in the semantic sense makes no difference on what the world is like and there is no metaphysical cost in such a move. Hence, it may be worth considering an alternative formulation of the conservativity requirement in the semantic sense: i.e.

If a theory S is deflationary, then S is conservative in the semantic sense over B. (1.4)

Cieśliński's paper shows that some disquotational theories are conservative in the proof-theoretic sense but not in the semantic sense. More precisely, he shows that only a recursively saturated model of PA can be expanded to a model of any of these disquotational truth theories. His paper is expected to shed more light on the distinction between the two formulations, 1.3 and 1.4, of the conservativity requirement, which has attracted less attention from philosophers than it should have, and is expected also to awake more technical interest in the problem of the semantic conservativity of theories of truth.

1.2.4.5 "New Constructions of Satisfaction Classes" by Ali Enayat and Albert Visser

Also concerned with conservativity, Enayat and Visser's paper "New Construction of Satisfaction Classes" presents a new proof of the conservativity of the axiomatic theory of a *full satisfaction class* over PA.

The modern definition of truth that we use today in model theory is usually credited to Tarski and Vaught (Tarski 1983; Tarski and Vaught 1956). They defined truth in

a model-theoretic structure $\mathfrak{M}$ via the definition of satisfaction for $\mathfrak{M}$. Given a structure $\mathfrak{M}$ for a language $\mathcal{L}$, we first define when each $\mathcal{L}$-formula ϕ is "satisfied" by a variable assignment α in $\mathfrak{M}$, which is a function from variables to the domain M of $\mathfrak{M}$, and thereby define that an $\mathcal{L}$-sentence is true in $\mathfrak{M}$ iff it is satisfied by at least one assignment α (or, equivalently, by all assignments α) in $\mathfrak{M}$. Hence, we may say that a theory of (Tarskian) satisfaction subsumes a theory of (Tarskian) truth. Enayat and Visser present a theory $\mathsf{PA}^{\mathsf{FS}}$ ("FS" for "full satisfaction class") of satisfaction over a base theory PA and in their paper prove that $\mathsf{PA}^{\mathsf{FS}}$ is conservative over PA for the language $\mathcal{L}_{\mathsf{PA}}$ of PA. This means that any arithmetical theorem of $\mathsf{PA}^{\mathsf{FS}}$ (i.e. an $\mathcal{L}_{\mathsf{PA}}$-sentence derivable from $\mathsf{PA}^{\mathsf{FS}}$) is already derivable in PA.

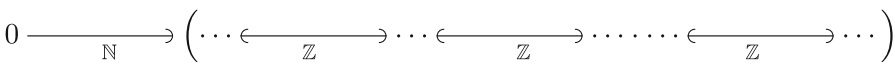
The theory $\mathsf{PA}^{\mathsf{FS}}$ is an axiomatic characterization of the Tarskian definition of model-theoretic satisfaction (for models of PA). It is formulated over the language $\mathcal{L}_{\mathsf{PA}}$ of PA plus a new binary predicate symbol S which takes a code or Gödel number of an $\mathcal{L}_{\mathsf{PA}}$-formula as its first argument and a variable assignment (of codes of variables to natural numbers) as its second argument, where $\mathsf{S}(x, y)$ expresses "a formula (coded by) x is satisfied by a variable assignment (coded by) y". For example, the $\mathsf{PA}^{\mathsf{FS}}$-axiom for negation is expressed as:

$$\forall x \forall y \forall u \big[\text{“}x \text{ is the code of the negation of the code } y \text{ of an } \mathcal{L}_{\mathsf{PA}}\text{-formula”} \wedge \text{“}u \text{ is a variable assignment”} \rightarrow \big(\mathsf{S}(x, u) \leftrightarrow \neg \mathsf{S}(y, u)\big)\big]. \tag{1.5}$$

For a model $\mathfrak{M}$ of PA with the domain M and for a binary relation $S \subset M \times M$, we say that S is a *full satisfaction class* for $\mathfrak{M}$ when $(\mathfrak{M}; S)$ is a model of $\mathsf{PA}^{\mathsf{FS}}$ in which S is interpreted by S. Precisely what Enayat and Visser prove in the present paper is that, for any model $\mathfrak{M}$ of PA, we can construct an elementary extension $\mathfrak{M}'$ of $\mathfrak{M}$ with the domain M' and a set $S \subset M' \times M'$ such that $(\mathfrak{M}'; S)$ is a model of $\mathsf{PA}^{\mathsf{FS}}$. This immediately entails the conservativity of $\mathsf{PA}^{\mathsf{FS}}$ over PA for $\mathcal{L}_{\mathsf{PA}}$ due to the completeness theorem.

Now, we know that Tarskian satisfaction can be defined for *any* structure. So, one might expect that $\mathsf{PA}^{\mathsf{FS}}$ is conservative over PA by reasoning in the following (wrong) way: since $\mathsf{PA}^{\mathsf{FS}}$ is an axiomatization of satisfaction that can be defined for every $\mathcal{L}_{\mathsf{PA}}$-structure, every model $\mathfrak{M}$ of PA can be expanded to a model $\mathfrak{M}^+$ of $\mathsf{PA}^{\mathsf{FS}}$ simply by interpreting S by the so-defined satisfaction for $\mathfrak{M}$, and we get the desired conservativity by the completeness theorem. It is true that $\mathsf{PA}^{\mathsf{FS}}$ is conservative over PA but this reasoning is fallacious; the proof of this conservativity is never that simple, and this fallacious reasoning is only valid when $\mathfrak{M}$ is the standard model of arithmetic.

By the compactness theorem, PA has a non-standard model and its non-standard part is ill-founded. In general, a non-standard model of PA has the following structure:



What we have here is many linear orderings of "non-standard numbers" *order-isomorphic to* $\mathbb{Z}$ topped up on the initial standard part that is order-isomorphic to $\mathbb{N}$.

In fact, a countable non-standard model of PA is order-isomorphic to $\mathbb{N} + \mathbb{Z} \cdot \mathbb{Q}$; see Kaye (1991, Chap. 6.2) for more details. In particular, each $\mathbb{Z}$-part of a non-standard model of PA has no end-points and is ill-founded with respect to the less-than relation $<$; and a non-standard number that lies on any $\mathbb{Z}$-part is greater than all standard numbers lying on the initial standard $\mathbb{N}$-part. Hence, we may informally say that non-standard numbers are "infinite" numbers. Since each non-standard $\mathbb{Z}$-part of a non-standard model of PA is ill-founded, the induction principle does not hold for these ill-founded $\mathbb{Z}$-parts. That is, for a non-standard model $\mathfrak{M}$ with the domain M, even if we have $\forall x \in M(\forall y \in M(y < x \rightarrow y \in X) \rightarrow x \in X)$ for $X \subset M$, we do not necessarily have $X = M$ (but $X = M$ holds under this assumption when X is $\mathcal{L}_{\mathsf{PA}}$-definable, since $\mathfrak{M}$ was assumed to be a model of PA). Another notable feature of non-standard models of PA is the "overspill" phenomenon. This is such that if an $\mathcal{L}_{\mathsf{PA}}$-definable property P is satisfied by unboundedly many standard numbers in the initial $\mathbb{N}$-part of a non-standard model $\mathfrak{M}$ of PA, then P is satisfied by some non-standard number of $\mathfrak{M}$ as well (or so to speak, the class $\{x \mid Px\}$ "spills over" the $\mathbb{N}$-part); see Kaye (1991, Chap. 6.1) for proof of this and more details. Consequently, a non-standard model $\mathfrak{M}$ of PA contains "non-standard $\mathcal{L}_{\mathsf{PA}}$-formulae". Precisely, for an $\mathcal{L}_{\mathsf{PA}}$-formula $\mathsf{Form}(x)$ that expresses "x is a code of $\mathcal{L}_{\mathsf{PA}}$-formula", $\mathfrak{M}$ contains a non-standard number a such that $\mathfrak{M} \models \mathsf{Form}(a)$; see Halbach (2010, § 8.3) for more detailed expositions. Similarly, a non-standard model of PA contains "non-standard variables", "non-standard syntactic complexities of formulae", etc.

Now, recall that the Tarskian definition of satisfaction for a structure $\mathfrak{M}$ is arrived at by recursion on the syntactic complexity of formulae (or the number of logical constants, etc.): we start the definition of satisfaction from the simplest formulae, i.e. atomic formulae; then, we define satisfaction for more and more complex formulae step by step using the definition of satisfaction already given for less complex formulae. Here, it is crucial that this definition applies directly to formulae and *not* to their "codes" in $\mathfrak{M}$. In contrast, however, the predicate S is interpreted in any model $\mathfrak{M}^+$ of $\mathsf{PA}^{\mathsf{FS}}$ as a relation over the domain M of $\mathfrak{M}^+$: the interpretation of S is a relation of M-elements satisfying a formula $\mathsf{Form}(x)$ in $\mathfrak{M}$ (i.e., M-elements *coding* $\mathcal{L}_{\mathsf{PA}}$-formulae in $\mathfrak{M}$) and M-elements representing variable assignments. In other words, the ordinary model-theoretic definition of satisfaction is for objects external to the structure $\mathfrak{M}$; whereas a full satisfaction class is to be defined for objects *in* the structure $\mathfrak{M}$. Hence, when $\mathfrak{M}$ is non-standard, the relation S may take a non-standard formula a whose syntactic complexity b is also non-standard. In such a case, an ordinary recursion or inductive definition of syntactical complexity (as an element of M) cannot reach the definition for a with complexity b, since b lies on an ill-founded $\mathbb{Z}$-part not accessible from the least element 0 by such a recursion process in a step-by-step manner.

This explains why the proof of the conservativity of $\mathsf{PA}^{\mathsf{FS}}$ is not as easy as it looks.[36] Furthermore, Lachlan (1981) showed that a non-standard model $\mathfrak{M}$ of PA only has a full satisfaction class when $\mathfrak{M}$ is *recursively saturated* (see Kaye 1991 for a definition of this). Therefore, not every model of PA has a full satisfaction class and therefore neither can they all be expanded to a model of $\mathsf{PA}^{\mathsf{FS}}$, since not all non-standard models of PA are recursively saturated.

As explained in Enayat and Visser's paper in more detail, the first conservativity proof for $\mathsf{PA}^{\mathsf{FS}}$ is given by Kotlarski et al. (1981). They show that every recursively saturated model of PA can be expanded to a model of $\mathsf{PA}^{\mathsf{FS}}$. So, since every model of PA has a recursively saturated elementary extension, this result yields the conservativity of $\mathsf{PA}^{\mathsf{FS}}$. To our knowledge, since that proof was offered by Kotlarski, Krajewski and Lachlan, there has been no other proof of this conservativity result (except for variants or extensions), and it is indeed quite technical and complicated. Enayat and Visser provide a simpler and more versatile proof of this conservativity result.[37]

1.2.5 Truth Without Paradox

The three papers in this chapter discuss solutions to the truth-theoretic paradoxes that escape sentences that are problematic in either their semantics or their syntax. Armour-Garb and Woodbridge argue for an approach in which some sentences employing the truth predicate, such as the Liar sentence, are to be seen as semantically defective. The authors defend the meaningless status of semantically problematic sentences, i.e. the fact that they lack wordly content, by means of a fictionalist account of truth-talk. Fictionalism in this context takes truth-talk to be part of a game of make-believe and the proper use of the truth predicate to be determined by the rules of that game. These rules establish certain worldly conditions as prescriptive for the pretenses displayed in (non-pathological) instances of truth-talk, so the talk thereby functions as an indirect means for specifying those conditions. The authors show how their account not only blocks the Liar paradox but also the reasoning underlying its revenge.

The view that some sentences that employ the truth predicate are semantically defective is also shared by Bonnay and Van Vugt; for them, however, semantic defectiveness amounts to lack of groundedness. In an attempt to make this idea

[36] The proof of the conservativity of the theory of satisfaction or Tarskian truth over set theory ZF is relatively easy. It was first model-theoretically shown by Krajewski (1976); and an elementary proof-theoretic proof can even be found in Fujimoto (2012).

[37] As Enayat and Visser state in their paper, the proof they present here still has some limitations in its application. For example, their proof assumes that the language of arithmetic is purely relational without any constant or function symbols. They announce in the paper that their techniques can be suitably modified for many other settings with functional languages and much weaker base theories. This immediately entails, for instance, that the theory $\mathsf{CT}\restriction$ of "compositional truth (with restricted induction)", also known as TC^{-} of "Tarskian inductive clauses (without expanded induction)", formulated in the standard functional language of arithmetic is also conservative over PA.

more precise, they compare two different ways of conceptualizing groundedness, as found in Kripke (1975) and Leitgeb (2005), and study the extent to which they are extensionally equivalent. In order to get a good grasp of the difference between the two approaches, Bonnay and Van Vugt take a closer look at Leitgeb's notion of conditional dependence, which they elucidate using the notion of groundedness according to the supervaluational scheme.

We saw earlier that in order to counter the so-called 'one concept objection' (according to which a hierarchy furnishes many concepts of truth when there is in fact only one), Glanzberg introduces the notion of stratification and argues that a concept may be stratified yet unique. In his contribution to this volume, Cantini proposes an axiomatization of the Tarskian hierarchy which essentially uses stratification in Quine's sense in order to represent a consistent theory of untyped truth. We should recall that the traditional theories of typed truth have been criticized for not being capable of representing the untyped ordinary notion of truth, and as a result various theories of untyped truth have been presented, such as Feferman's KF and Cantini's own VF. Cantini adds a new kind of theory of untyped truth to the existing variety of such theories; it is a theory of stratified—hence, not generally self-referential—untyped truth.

1.2.5.1 "Truth, Pretense and the Liar Paradox" by Bradley Armour-Garb and James A. Woodbridge

In their contribution Armour-Garb and Woodbridge propose a new fictionalist framework to articulate deflationism about truth. They argue that their construal of deflationism allows for a way out of the truth-theoretic paradoxes. According to the brand of fictionalism that the authors defend—pretense-involving fictionalism, as they call it— truth-talk always invokes a background game of make-believe in its functioning, in virtue of which these sentences serve as an indirect means for specifying (as obtaining or not obtaining) the worldly conditions that the game's rules establish as prescriptive for the pretenses that the sentences display. These constitute the meaning-conditions specified by the sentences. Thus, for instance, our truth-involving language game compels us to pretend that expressions such as 'is true' are descriptive predicates—which they no more are than children pretending to be cowboys in the courtyard really are cowboys—and also that the pretenses displayed in a utterance of 'It is true that snow is white' are prescribed if and only if snow is white, and so on. The rules continue in a similar manner for other classical conditions that deflationists take to govern the use of the truth predicate. The authors claim there are two main benefits of their approach over other deflationist frameworks. The first is methodological: it allows for unification of the deflationist treatment of truth-talk with widely used fictionalist strategies in other areas of philosophy. More specifically, in this paper the authors argue for a second benefit: that their pretense account of truth is immune to the Liar paradox and its revenge. This is because, the authors argue, their account allows for a successful version of a "meaningless" strategy against these paradoxes.

Regarding the simple Liar sentence, the rough idea of the strategy is as follows. The Liar sentence is an instance of truth-talk, and as a case of pretense talk, any

meaning it had would be constituted by whatever worldly conditions the rules of the pretense establish as prescriptive for the pretenses it displays. However, the Liar sentence manifests a kind of 'ungroundedness' because the rules of the pretense do not make any worldly conditions prescriptive for the pretenses it displays, and so the Liar sentence specifies no meaning-conditions. In this way, sentences like the Liar suffer a kind of semantic defectiveness that the authors go on to partially characterize in their essay. Furthermore, because such sentences do not specify any meaning-conditions, there are no conditions that are prescriptive for the pretenses invoked in asserting that the Liar sentence is true, or that it is false, or even that it is not true. All such instances of truth-talk are also semantically defective. It follows that sentences like the Liar are not truth-evaluable at all. What about the revenge? It is well known that many solutions to the Liar paradox that accept partitioning the domain of sentences into true sentences, false sentences, and sentences with a third kind of truth value, I, are generally vulnerable to a revenge paradox, by way of reasoning involving a sentence which says of itself that it is false or I. Let λ be the sentence: λ is false or λ is I. The usual reasoning is: (1) λ cannot be true, since this would imply that it is false or I (both of which contradict the hypothesis that λ is true). (2) If λ is false, then its first disjunct is true, hence λ is true after all—again contradicting the hypothesis. (3) Then finally, if we hold λ to be I, λ must be true in virtue of its second disjunct, which now contradicts its being I—and we are caught in the revenge paradox.

In the authors' framework, no third truth-value is involved, but one could construct a corresponding tentative revenge sentence in the form of a sentence λ which says that λ *is not true or* λ *is semantically defective*. The authors argue that this sentence is semantically defective. Moreover, and this is where their work on the meaning-conditions of truth-talk pays off, it is argued that even though the sentence 'λ is semantically defective' is true, that does not imply that the sentence 'λ is not true or λ is semantically defective' (i.e., λ itself) is true. This is because: first, λ being devoid of content means that its first disjunct is an instance of truth-talk that (as in the simple Liar case) fails to specify any meaning-conditions; and second, disjoining a meaningless sentence with a true sentence results in a meaningless whole, in virtue, roughly speaking, of the compositional features of meaning constitution. Since a meaningless sentence cannot be a consequence of anything, it follows that λ being semantically defective blocks the inference from the claim that it is semantically defective to λ itself. Thus, the air of paradox is dispelled as, specifically, the semantic defectiveness of the problematic sentence λ no longer implies its truth. Carrying on with their earlier work on meaning and understanding in connection with liars, the pretense-involving fictionalism defended by the authors here contributes to giving new bite to "meaningless strategies" against the paradox.

1.2.5.2 "Groundedness, Truth and Dependence" by Denis Bonnay and Floris Tijmen Van Vugt

Gupta and Belnap (1993) argued that in order to make a correct diagnosis of the truth-theoretic paradoxes one should not focus on the paradoxical sentences themselves,

but instead try to understand the ordinary behaviour of the truth predicate. In the same spirit, Bonnay and van Vugt are interested here not in determining which sentences are pathological, but in characterizing unproblematic ones. The aim of their paper is to compare two prominent characterizations of non-pathological sentences due to Kripke (1975) and Leitgeb (2005). Both characterizations share the basic intuition that unproblematic sentences are those that are grounded in the world, i.e. those whose truth value ultimately depends on what the world is like. But Kripke and Leitgeb propose two different accounts of groundedness that are not extensionally equivalent and here Bonnay and van Vugt wish to identify the parameters responsible for their divergence.

Kripke follows an indirect route in order to determine the collection of grounded sentences, in the sense that he first defines the extension of the truth predicate via a fixed-point construction and then he defines grounded sentences as those that have a truth value at the least fixed point. Of course, which sentences are grounded also depends on the scheme of evaluation used. Leitgeb, in contrast, uses a direct route; he determines the class of grounded sentences as a fixed point of a construction that directly identifies sentences whose truth value does not depend on the truth predicate. In what follows, J^K denotes the Kripke jump for the strong Kleene scheme of evaluation; then K-grounded sentences are those grounded in Kripke's sense according to the least fixed point of J^K. J^V, and V-grounded are the corresponding notions for the supervaluational scheme. L-grounded sentences are those grounded according to Leitgeb's characterization.

L-groundedness does not imply K-groundedness; nor vice versa. As an example of the difference between L-grounded and K-grounded sentences, the authors consider the sentence $Tr\ulcorner 2+2=4\urcorner \vee \lambda$, where λ is the Liar sentence. As $2+2=4$ is true, $Tr\ulcorner 2+2=4\urcorner$ is also true, which makes $Tr\ulcorner 2+2=4\urcorner \vee \lambda$ a K-grounded sentence. However, according to Leitgeb's definition of dependence, the sentence $Tr\ulcorner 2+2=4\urcorner \vee \lambda$ depends both on $2+2=4$ and on λ, hence it is L-ungrounded; although $2+2=4 \vee \lambda$ does not depend on any other sentence and so it is L-grounded. This is a result that the authors find counterintuitive and they explore ways to avoid this. They consider Leitgeb's notion of conditional dependence, according to which sentences that are declared grounded in one step of the construction depend on the grounded sentences in the previous step and on the partition of those into true and false sentences. The main result of the paper, which contributes to a better understanding of the connection between these two notions of groundedness, is that the set of grounded sentences for conditional dependence coincides with the V-grounded sentences.

1.2.5.3 "On Stratified Truth" by Andrea Cantini

Cantini's contribution is motivated by the problem of "finding a consistent axiomatization of the Tarskian hierarchy, where stratification is understood in Quine's sense", which he accredits to Feferman. Let us start by explaining what "stratification in Quine's sense" means formally.

Russell's and other set-theoretic paradoxes suggest that "circularity" must be somehow restricted in set theory for the sake of consistency. This view led to the

idea of type hierarchy, according to which, each set is assigned a "type" and may contain only sets of lower types. "Circularity" is thereby excluded by stipulating that the types form a well-founded hierarchy; for then the type of any set a cannot be lower than itself.

A natural formal implementation of this idea is the so-called theory of types. In the theory of types, each item of vocabulary of a language is syntactically assigned types. For instance, each variable is indexed by a natural number $i < \mathbb{N}$, indicated by writing v^i, where the hierarchy of the types is given by the less-then relation of natural numbers, and an expression $v^i \in v^j$ is syntactically well-formed only when $j = i+1$. Then each compound formula is uniquely assigned a type in a straightforward manner by taking the maximum of the types that occur in the formula, and the axiom schema of comprehension is reformulated as

$$\exists x^{i+1} \forall z^i (z^i \in x^{i+1} \leftrightarrow \phi(z^i)), \text{for a formula } \phi \text{ of type } i+1.$$

Quine's New Foundation, NF, uses a different method to formally implement the idea of type distinction. In NF, each item of vocabulary, such as a variable and the membership relation $\in$, is *not* indexed by any type *at the syntax level*; the language of NF is an ordinary first-order language of set theory with a single sort. However, instead of syntactical type distinctions, NF restricts the axiom schema of comprehension to formulae that *could* be typed by natural numbers; those potentially "typable" formulae are said to be *stratified* in this context. For example, $x \in x$ is not stratified since there is no possible type assignment in which the type of x is higher than that of x. In contrast, $x \in y$ is stratified (provided that $x \neq y$) since we can assign any natural number n to x and then assign $n+1$ to y; so NF reformulates the axiom's schema of comprehension as follows:

$$\exists x \forall y (x \in y \leftrightarrow \phi(x)), \text{for a stratified (i.e. typable by natural numbers) formula } \phi.$$

At first glance, the difference between the two formal implementations of the idea of type distinction looks superficial, and one might expect that NF is essentially the same as the theory of types. This is not the case, however. For a typical example, the universal set provably exists in NF since $x = x$ is obviously stratified; in contrast, there is no such set in the ordinary theory of types. Interestingly, it turned out that NF is quite a complex theory and, in fact, there is no consistency proof for it yet.

Truth predicates are applied to terms that refer to sentences, and the Liar paradox is (usually diagnosed to be) caused by a self-referential application of a truth predicate; specifically, it is caused by a sentence in which a truth predicate is applied to a term that refers to the sentence itself. Given this, the Liar paradox and its cousins suggest, in essentially the same way as Russell's paradox suggests for set theory, that "self-reference" must be somehow restricted in theories of truth in order to avoid inconsistency. The argument of Tarski's undefinability theorem, which is a formalization of the Liar paradox within formal systems, tells us that the following so-called T-schema

$$T(\ulcorner \sigma \urcorner) \leftrightarrow \sigma, \text{for all sentences } \sigma,$$

cannot be consistently sustained if we allow σ to contain the truth predicate T. This naturally suggests the view that application of a truth predicate should be restricted to sentences that do not contain that predicate.

This view gave rise to the theory of typed (or ramified) truth. The language of a theory of typed truth contains a stock of more-than-one truth predicates each of which is indexed by its type, i, and the application of the truth predicate of type i is restricted to sentences that only contain truth predicates of types lower than i. As in the case of the theory of types, each formula of a theory of typed truth is uniquely assigned its type. So the truth theoretic axioms such as the T-schema are reformulated to:

$$T_i(\ulcorner \sigma \urcorner) \leftrightarrow \sigma, \text{ for a sentence } \sigma \text{ of type lower than } i.$$

Now we can see a clear analogy between the theory of types and the theory of typed truth in their remedy for the paradoxes.

Cantini's new theory SFT takes an alternative route to consistently restrict "self-referential" applications of a truth predicate in the same way as Quine's NF does for consistently restricting "circularity". Cantini adopts a language with a single universal truth predicate T without any index or "typing", but introduces the notion of *stratified* formulae of truth theory in the same spirit as Quine's notion of stratified formulae of set theory: a stratified formula is a formula in which we could suitably type the occurrences of terms and the truth predicate T by natural numbers; see Definition 1.3 of the paper. Thereby, Cantini restricts the T-schema as well as the other truth-theoretic principles to stratified sentences (more generally, stratified formulae); e.g.

$$T(\ulcorner \sigma \urcorner) \leftrightarrow \sigma, \text{ for a stratified sentence } \sigma.$$

For instance, the Liar sentence λ has the property $\ulcorner \lambda \urcorner = \ulcorner \neg T \ulcorner \lambda \urcorner \urcorner$, but a term such as $\ulcorner \lambda \urcorner$ is not stratified and properly excluded from the range of application of the truth predicate in the stratified T-schema. Cantini then proceeds to provide a proof of the relative consistency of his SFT to NF. Hence, if NF is consistent, SFT is consistent; although we do not yet know whether NF is consistent or not.

1.2.6 Inferentialism and the Revisionary Approach

The title of the previous chapter, 'Truth without Paradox', fits the papers collected here well too, since they present work primarily motivated by the goal of evading the truth-theoretic paradoxes. The reason for grouping these papers into a separate section is to emphasize their interest in the inferential substrate underlying the derivation of the paradoxes, whether it is to do solely with principles governing truth or it extends to the logic. All the papers can be said to follow a revisionary approach; albeit not in the same sense. Theories of truth are usually called revisionary if they revise the underlying logic to avoid compromising naive properties of truth. A well-known example of such an approach can be found in Priest's contribution at the end of this chapter where a paraconsistent logic is proposed to replace classical logic. However,

according to Murzi and Shapiro, as well as Zardini, such revisions are only partly successful. Those authors take an even more radical approach in proposing a revision of the structural rules of the underlying logic; and in fact both of their papers draw our attention to the rule of contraction. Also on the substructural level, Cobreros, Egré, Ripley and van Rooij challenge the rule of transitivity, though only for reasoning that involves truth. This focus on truth-related reasoning[38] is shared by Read who is, however, primarily concerned with the (in)validity of specific principles of truth, as opposed to the logic.

Read exploits a solution to the Liar paradox due to the medieval philosopher Bradwardine, according to which the Liar sentence comes out false. The solution is based on Burley's semantics of signification and truth. Read crucially explores the semantics and logic underlying this solution to the paradox which he shows to have some attractive features. In particular, and unlike the following papers, the logic itself remains intact, while some problematic truth principles are proved by Read not to follow from the semantics; in particular, T-OUT (i.e. $T^{\ulcorner}\phi^{\urcorner} \rightarrow \phi$) and commutation of truth with negation and conditional. This gives a principled way for motivating a consistent notion of truth.

Inspired by their treatment of vague predicates, Cobreros, Egré, Ripley and van Rooij propose a novel consequence relation, their so-called strict-to-tolerant consequence relation, which is permissive in that it allows for tolerantly asserted conclusions (conclusions that take the value 1 or 1/2) from strictly asserted premises (premises that take the value 1). This consequence relation basically blocks the Liar paradox by invalidating transitivity for inference steps whose conclusions are only tolerantly accepted; while, given its permissive character, it preserves transparency for truth. The authors explain why dropping transitivity for some inferences that involve truth is not as implausible as one may think at first. As with Read's paper, the underlying logic here is not affected in the absence of the truth predicate, since strict-to-tolerant and classical consequence coincide in this case.[39]

From permissive consequence, Murzi and Shapiro take us back to the traditional intuitive notion of validity as truth preservation (VTP) which is often dismissed by revisionary theorists of truth. The authors believe that just as the revisionary theorist is interested in guarding the naive properties of truth, importantly the transparency (or intersubstitutivity) of truth, naive properties of validity should also be safeguarded. They rehearse the usual reasons for repudiating VTP and argue that these are based on a solution to the semantic paradoxes that is, however, problematic. Showing then that rejecting the rule of contraction offers a better solution to the paradoxes—a rule the authors believe in any case to be in tension with naive principles governing the intuitive notion of validity—the reasons for dismissing VTP fall through.

---

[38] Recall that the term 'revision' is also used in the sense of applying only to truth by Halbach and Horsten in this volume.

[39] It is interesting to observe that transitivity is also what is at stake in the medieval sophism that Read discusses at the beginning of his paper as the starting point for developing an alternative semantics for signification and truth.

Zardini is on the same page as the previous two papers in considering the transparency of truth a necessary requirement for an adequate theory of truth. His endorsement of a contraction-free logic comes from comparing it with paracomplete and paraconsistent theories in their capacity to provide a solution to three paradoxical arguments that he presents. Zardini shows that not only is his proposal able, unlike the other theories, to deal with all three arguments, but that it does so by offering a unified solution to them. This conclusion is then philosophically strengthened even further by metaphysical motivation for rejecting the rule of contraction.

All the papers in this chapter endorse a unified approach to the semantic paradoxes. We find, for example, that Read, Murzi and Shapiro, and Zardini all discuss Curry's paradox. For Read, Curry's paradox plays a crucial role in improving Bradwardine's logic so that there is a principled reason (other than simply evading the Liar) for excluding commutation of truth not only with negation, but also with implication. Murzi and Shapiro, crucially base their support for a contraction-free based logic on the failure of alternative proposals to deal with Curry's paradox; as does Zardini. Finally, Cobreros, Egré, Ripley and van Rooij extend the idea of a unifying approach from the semantic paradoxes to the paradoxes of vagueness.

It is, therefore, fitting to conclude this chapter, and with it the volume, with Priest's paper which is a systematic expression of the idea of a uniform solution to the paradoxes. Priest revisits his Inclosure Schema which provides a general form underlying self-referential paradoxes. He tests and strengthens the validity of this schema by showing that a new intensional paradox, Kripke's thought paradox, falls under it. Subsequently, in accordance with his so-called Principle of Uniform Solution, he shows that dialetheism, which he has earlier defended as the solution to the Liar paradox, applies in this case too.

1.2.6.1 "Truth, Signification and Paradox" by Stephen Read

In his contribution Read engages in the discussion of Bradwardine's solution to the Liar paradox; one of the medieval solutions to the semantic paradoxes that is the focus of renewed interest today. The topic is introduced with a sophism that attracted some attention in the middle ages: "If I say that you are an ass, I say that you are an animal. And if I say that you are an animal I say something true. Therefore, if I say you are an ass, I say something true". The conclusion seems obviously false; so, if the premises are true, then the validity of the transitivity of the conditional (suffixing, in Read's terminology) is threatened. Read surveys several authors who all respond to the sophism by distinguishing two notions of "saying that". One corresponds more or less to the literal notion of meaning: the second premise is true because if I say that you are an animal I say something true; but the first premise is false because if I say that you are an ass I do not literally say that you are an animal. According to the second notion, when I say that you are an ass, I also say all the consequences of that statement; so the first premise is true, but then the second premise is false because when I say that you are an animal I do not always say something true (for example, if I say that you are an animal by means of saying, falsely, that you are an ass).

Read focuses on Burley's account of these ideas. Burley distinguishes four notions of proposition: the written proposition, the spoken proposition, the mental

proposition and the real proposition. The last two are a composition (or division) of concepts or real things, respectively. A written or spoken proposition signifies a mental proposition which is true when it corresponds to a true real proposition, i.e. when it composes concepts that stand for things which are really united in reality or divides concepts that correspond to things really divided in reality. The two key aspects of Burley's semantics are that signification is closed under consequence (hence, if I say that you are an ass, I also say that you are an animal), and that truth requires that all the things I say are in reality as signified (so that sometimes I may be saying something false when I say that you are an animal, because I may say so by saying that you are an ass).

The second section of the paper explains how these two theses are used by Bradwardine to develop an original solution to the Liar paradox. In the same way that a proposition that signifies that you are an ass also signifies that you are an animal, a proposition that signifies that itself is not true, also signifies that itself is true, because this follows from it. Hence the Liar proposition signifies both that it is true and that it is false and, therefore, the Liar is simply false, because not everything that the Liar signifies is true. Read explains in detail the derivation that Bradwardine gives of this result and also proves that from his postulates it follows that every sentence signifies its own truth, as was defended by several medieval logicians.

The last section of the paper studies the logic implicit in Bradwardine's solution to the paradoxes. He accepts half of the T-schema (T-OUT: if p is true, then p) but not the converse, since from the fact that p obtains, it does not follow that everything that p signifies obtains. Read then focuses on the analysis of the compositional principles of truth that fail for Bradwardine's solution; a problem shared with other solutions that deny T-IN (such as Kripke 1975 and Maudlin 2004). Bradwardine just adds the compositional principles for disjunction and conjunction as axioms (a conjunction is true iff each part is true, and it is false iff one of its parts is false; analogously for disjunction). But that does not satisfy Read, because, for instance, if we added similar compositional principles for negation or the conditional, we would arrive at paradoxes (the Liar and Curry paradoxes, respectively). Hence, how do we know that adding the compositional principles for conjunction and disjunction do not create new paradoxes? (Think of simple paradoxical sentences such as '1 + 1 = 2 and this whole sentence is false'.) Read offers an argument on behalf of Bradwardine showing that the compositional principles for conjunction and disjunction can be proved from the basic principles of logic that Bradwardine accepts. He defends this solution as attractive because it "preserves those truth principles which are unaffected by the paradoxes, without sacrificing any logical principles".

1.2.6.2 "Vagueness, Truth and Permissive Consequence" by Pablo Cobreros, Paul Egré, David Ripley and Robert van Rooij

In their contribution, Cobreros, Égré, Ripley and van Rooij adopt a non-standard notion of logical consequence in order to provide an adequate semantics for vague predicates as well as for the truth predicate. A predicate is vague when it satisfies the principle of *tolerance*: sufficiently small variations in two objects cannot make

a difference in the application of the predicate. For instance, if one person is only 1 cm shorter than another person, then either both of them are tall or neither of them is. The principle of tolerance, together with classical logic, produces the sorites paradox: consider a series of people from someone who is clearly tall to someone who is clearly not tall, such that any person immediately following another person in the series is just 1 cm shorter. Given that the first person in the series is tall, by a repeated application of the principle of tolerance it follows that the last person in the series is also tall, contradicting the initial hypothesis that the last person is not tall. The analogous key semantic feature governing the truth predicate is the *transparency* principle which states that a sentence 'A' is intersubstitutable with 'A is true' in all extensional contexts without any cost for the validity of arguments. Transparency together with classical logic produce the Liar paradox and its kin.

There have been well-known attempts in the literature to solve these paradoxes by the use of three-valued logics that admit gappy sentences (i.e. whose truth value is neither true nor false) in order to deal with borderline cases of vague predicates or the truth predicate (Kripke 1975 and Tye 1994). Those solutions have been subjected to close scrutiny and often criticized (see, for example, Keefe 2000 and Gupta and Belnap 1993). In this paper the authors try to overcome the limitations of the standard solutions not by introducing a new three-valued scheme of interpretation (they follow the standard strong Kleene interpretation), but by modifying the definition of logical consequence. The new definition is based on two modes of assertion: *strict* and *tolerant*. In a three-valued setting, a sentence is strictly assertible provided it takes the value 1, and tolerantly assertible provided it does not take the value 0. Given the strict and tolerant standards of assertion, the authors explore the consequence relations that arise from varying these standards for premises and conclusion. They show in particular how to combine the two modes to get a relation of permissive or st-consequence consequence (strict to tolerant), which requires premises to be asserted strictly and the conclusion only tolerantly. This consequence relation combines features of both Kleene's Strong Logic (K3) and of its dual, the Logic of Paradox (LP).

The notion of permissible consequence has some desirable properties that, according to the authors, make it appropriate for providing a good solution to the paradoxes of vagueness and truth. For a language that does not contain vague predicates or the truth predicate, st-consequence coincides with that of classical logic; which also means that it has a well-behaved conditional that satisfies modus ponens and the deduction theorem. In the presence of vague predicates or the truth predicate, it also satisfies the tolerance and transparency principles. So what prevents the sorites and Liar arguments from going through is the fact that st-consequence lacks *transitivity*. Although for every object a_n in a sorites series, it follows that: if a_n is tall, then a_{n-1} is also tall, we cannot chain all these reasoning steps together to conclude that if the first object in the series is tall, the last one is too. As an example of non-transitivity in the case of truth, even though any sentence has as an st-consequence the Liar, and the Liar has as an st-consequence any sentence, explosion does not follow.

Alongside applying st-consequence to the cases of vagueness and truth, the authors discuss two concerns that are naturally raised by their proposal. On the one hand, since transitivity is a basic structural rule, it is difficult to justify how we

can do ordinary reasoning without it[40]. On the other hand, they address the standard concerns of how to solve issues of higher-order vagueness or strengthened Liar paradoxes.

With respect to the first problem, the distinction between strict and tolerant assertion comes to rescue. Transitivity still holds when we go from strictly accepted premises to strictly accepted conclusions, but fails when we get only tolerated conclusions that cannot subsequently be used as strictly accepted premises. This means that in the absence of vagueness or semantic predicates, we can safely reason transitively.

With respect to the second problem, they consider revenge paradoxes expressed with the determinateness operator: we want to say that, if an object is a borderline case of tallness, then it is neither determinately tall nor determinately not tall; and we also want to say that the Liar paradox is neither determinately true nor determinately false. In a three-valued logic, the natural determinateness operator is an operator D such that DA is true when A is a true sentence and false when A is either a false or a gappy sentence. But then if a sentence A says that an object is a borderline case of tallness, then A is gappy and, given the definition of D, the sentence $D(\neg DA \wedge \neg D\neg A)$ is true. This means that it is determinately the case that the object is a borderline case of tallness. Hence, one cannot express the existence of second-order vagueness in the language (i.e. the existence of objects that are not determinately tall but also not determinately borderline cases of tallness). In the case of truth, if D belongs to the language, then the sentence that says of itself that it is not determinately true is a new paradox that the three-valued semantics cannot consistently evaluate. The authors of the paper explore two strategies for coping with these revenge paradoxes and claim that their theory is compatible with both. The first strategy, which is the one they find most congenial to their theory, is to argue that those operators should not be included in the object language. The second strategy consists of modifying the semantics to include some versions of determinateness operators.

1.2.6.3 "Validity and Truth-Preservation" by Julien Murzi and Lionel Shapiro

In their contribution, Murzi and Shapiro address one of the unpleasant consequences of revisionary approaches to paradox, which is that some naive principles governing the intuitive notion of validity are invalidated. The authors consider the standard definition of validity as truth-preservation, so-called VTP: 'If an argument is valid, then, if all its premises are true, then its conclusion is also true'. Validity relies on truth according to VTP, so it is to be expected that attempts to secure a consistent notion of truth may have consequences for VTP. The authors first rehearse the reasons why VTP has come to be unpopular among revisionary theorists. First, invalidating

[40] Recall Feferman's relevant dictum which in their contribution to this volume Halbach and Horsten present as one of the desiderata for type-free theories of truth.

VTP is a corollary of the standard revisionary approach to the semantic paradoxes of truth in that revisionary approaches typically aim to preserve naive properties of truth (crucially, the T-schema and intersubstitutivity) and propose a conditional to replace the so-called detaching conditional (i.e. one that satisfies MP), in order to deal with the Liar paradox. In order to secure a uniform approach to the semantic paradoxes, this new conditional is typically chosen to be such that it also blocks the derivation of the c-Curry paradox (c for conditional). This is achieved by a conditional that does not satisfy the $I \rightarrow$ rule, thereby invalidating the law of contraction, which is widely held responsible for the c-Curry paradox. But without $I \rightarrow$, VTP is also invalidated, for example, in its simplest reconstruction offered by Field; the so-called Validity Argument (Field 2008).

Independent arguments have also been put forward to show that VTP cannot be consistently asserted in the object language. The authors believe that VTP is a factive statement, and that a semantic theory should be able to affirm what we know to be true. They thus counterpose their own independent arguments in defence of VTP. These come down to two main points. First, the authors believe that a 'naive view of semantic properties' is in the spirit of the revisionary approach to semantic paradox in general and not only for truth. To make the 'naive view of validity' precise, the authors list two principles that underwrite it: the so-called VP and VD, which resemble the necessitation rule and the **T** axiom for the necessity operator respectively. Their claim is that just as dealing with the Liar paradox does not call for revision of the semantic properties of truth by, for instance, invalidating the T-schema, so the revisionist should not abolish the naive VP and VD either when seeking a way out of the paradoxes of validity. Second, the standard revisionary way of dealing with the Curry paradox, rejecting the operational rule of $I \rightarrow$, does not apply in the case of the v-Curry paradox (v for validity), since the rule is not used in the derivation. In contrast, the structural rule of contraction *is* used and since this paradox of validity is as genuine a paradox as c-Curry is, a uniform solution to both versions of Curry's paradox naturally calls for dispensing with contraction.

The authors proceed to inquire into the consequences of their proposal; that is, the consequences that the rejection of contraction has on the main challenges to VTP that they have identified. They show that these challenges to VTP rely on the way premise aggregation is represented for multiple-premise arguments. Moreover, they show that rejecting contraction opens up different non-classical options to represent premise aggregation (as in multi-based logic with multiplicative conjunction or dual-bunching logic). The authors then offer a detailed analysis of where the arguments underlying the challenges to VTP break down with respect to different substructural choices. Rather than arguing in favour of one such choice, the authors' aim is to establish that VTP is not incompatible with a revisionary approach to paradox; especially since rejecting contraction is no obstacle to supporting a naive theory of truth.

1.2.6.4 "Getting One for Two, or the Contractors' Bad Deal. Towards a Unified Solution to the Semantic Paradoxes" by Elia Zardini

In his contribution, Zardini proposes a new solution to the Liar paradox based on a substructural logic (see also Zardini 2011). He takes the transparency of the truth predicate to be a requirement of any acceptable theory of truth and argues that the basic law that has to go in order to restore the consistency of a truth predicate in a self-referential language is the structural rule of contraction. In its most basic form, that rule says that, given sentences ϕ, ψ, if $\phi, \phi \vdash \psi$, then $\phi \vdash \psi$. The version of contraction-free transparent theory of truth Zardini uses is called IKT and is presented in sequent calculus style. The paper compares the paracomplete and paraconsistent treatment of the paradoxes (that reject LEM—the Law of Excluded Middle—and LNC—the Law of Non-Contradiction—respectively) with IKT, and argues that the latter offers an important advantage over its competitors: it provides a unified solution to different paradoxical arguments, while paracomplete and paraconsistent solutions need to concoct different modes of justification to solve different paradoxical arguments.

The bulk of the paper is devoted to presenting three paradoxical arguments followed by a detailed analysis of the diagnosis that the different theories could give of them. The first two arguments use the Liar, i.e. a sentence λ identical to $\neg T(\ulcorner \lambda \urcorner)$, to produce unacceptable (even for a dialetheist) conclusions. In the first argument both LEM and LNC are used, so paracomplete and paraconsistent solutions have a principled way to reject it. The second argument, however, does not use LEM. This would force the paracomplete theorist to reject another rule, typically the metarule of the single-premise reduction theorem (if $\phi \vdash f$, then $t \vdash \neg\phi$, where t expresses the conjunction of all logical truths and f the disjunction of all logical falsehoods). Zardini gives a detailed analysis of the justification for this rule to show that its rejection does not follow from the denial of LEM. A parallel of this dialectics is presented, offering a version of the second argument that does not use LNC and that would force the dialetheist to reject the metarule of the single-conclusion demonstration theorem (if $t \vdash \phi$, then $\neg\phi \vdash f$) for reasons that do not follow from the denial of LNC. The third form of argument analysed in the paper is Curry's paradox, which uses neither LEM nor LNC. In this case, paracomplete and paraconsistent theorists would have to reject the single-premise deduction theorem (if $\phi \vdash \psi$, then $t \vdash \phi \rightarrow \psi$), but only for reasons that do not follow from the denial of either LEM or LNC. In contrast to paracomplete and paraconsistent theories, IKT accepts as valid all the principles dismissed by them and offers a unified solution: reject contraction, which is used in all three arguments. Of course this advantage by itself does not show why contraction fails. Although this question is not tackled here, the paper ends by giving a few hints of an answer that relies on a picture of metaphysical reality as unstable. Zardini considers that contraction fails for sentences that express unstable states of affairs, meaning states of affairs that lead to consequences with which they need not co-obtain.

1.2.6.5 "Kripke's Thought-Paradox and the 5th Antinomy" by Graham Priest

As a consequence of the central role that the Liar paradox occupies in the philosophy of truth, the reasoning form of the paradox has also become an object of study in its own right, independent of underlying theories. Since it can be classified as a self-referential paradox, a way to further our understanding of it is to study whether there are features shared by self-referential paradoxes which license a common description of them. The so-called Inclosure Schema has been proposed by Priest (2002) as such a description. The schema is a reworking of Russell's description of set-theoretic paradoxes given in 1905: 'On some difficulties in the theory of transfinite numbers and order types' (Russell 1906). Priest generalizes Russell's description to self-referential paradoxes (which, for example, include the Liar and König's paradox). He is careful to note that not all reasoning that shares this underlying form should be considered as paradoxical (that is, the schema should not be read as laying out sufficient conditions); this can already be seen in his discussion of the Barber paradox in his (2002), but Curry's paradox also presents a different challenge. What the extra conditions are for such self-referential reasoning to count as paradoxical is still under discussion, and as a consequence, so is the meta-theoretic status of the schema (for instance, whether it can be seen as playing a heuristic role).

Priest's interest in having a general schema that delineates a class of paradoxes is strongly linked to his belief that paradoxical arguments that share the same form call for a uniform solution. He introduced this idea in his (1994) under the name of the 'Principle of Uniform Solution'. In his contribution to this volume, both the Inclosure Schema and the Principle of Uniform Solution are put to the test by considering a new intensional self-referential paradox, namely, Kripke's thought paradox. In the first part of the paper Priest shows how the Inclosure Schema can accommodate that paradox and in the second part he argues for a dialetheist solution to it. Since the same solution can be given for the Liar paradox, this paper contributes, on the one hand, to a uniform representation and solution of self-referential paradoxes; as well as providing an additional argument for dialetheism, on the other. We saw that the preceding papers depart from the specific paraconsistent approach as the solution to the paradoxes. Yet, Priest's work constitutes a systematic expression of the philosophical idea of unification that drives most of the work in this area of revisionary approaches to truth.

Acknowledgements We would like to thank Springer and especially Shahid Rahman for hosting this collection of papers within the series 'Logic, Epistemology and the Unity of Science'; we are thankful to the editorial office of Springer for all their work, and especially Christi Lue and Rajdeep Crest Roy. We are most grateful to the authors of the papers in this volume for honouring us with their contributions, as well as their patience, cooperation and trust while we have been preparing this volume. Finally, we are grateful to all those who have acted as anonymous referees for their careful reading and valuable feedback.

Theodora Achourioti would like to thank Peter van Ormondt for his help with the organization of the conference 'Truth be told' (Amsterdam, 23–25 March 2011). She would also like to acknowledge the generous financial and technical support offered by: the NWO open competition

project ‘The Origins of Truth and the Origins of the Sentence’; the Institute for Logic, Language and Computation; the Evert Willem Beth Foundation; the NWO VICI project ‘Unsupervised Learning with the DOP Model’; and the Philosophy Department of the University of Amsterdam.

Kentaro Fujimoto would like to thank Volker Halbach as co-organizer of the “Axiomatic Theories of Truth” conference (Oxford, 19–20 September 2011) as well as the other members of the AHRC project “Inexpressibility and Reflection in the Formal Sciences” for their cooperation; and also to acknowledge financial support for the conference by the Arts and Humanities Research Council (AH/H039791/1).

Henri Galinon would like to thank Pierre Wagner and Denis Bonnay as co-organizers of the “Truth at Work” conference (20–23 June 2011), the members of the Institut d’Histoire et de Philosophie des Sciences et des Techniques for their help in organizing the conference, as well as Sebastien Gandon and the PHIER for their support during the preparation of this volume. The organization of the conference was financially supported by the ANR-funded project ‘Logiscience’ managed by Pierre Wagner at the IHPST.

José Martínez would like to thank Genoveva Martí and Sergi Oms as co-organizers of the “BW7 Conference: Paradoxes of Truth and Denotation” (14–16 June 2011) as well as the other members of the Logos Research Group for their cooperation. He would also like to acknowledge funding received from the Spanish *Ministerio de Economía y Competitividad* for project FFI2010-11447-E, R+D Project FFI2011-25626 (Reference, Self-reference and Empirical Data) and the Consolider Ingenio Program, CSD2009-00056, (Perspectival Thoughts and Facts); and from the *Generalitat de Catalunya* for project 2010ARCS10160.

References

Aczel, P. (1980). Frege structures and the notions of proposition, truth and set. In J. Barwise, H. Keisler, & K. Kunen (Eds.), *The Kleene symposium* (pp. 31–59). Amsterdam: North-Holland.

Barendregt, H. (1984). *The Lambda calculus, its syntax and semantics (studies in logic and the foundations of mathematics, volume 103)*. Amsterdam: Elsevier.

Beeson, M. (1985). *Foundations of constructive mathematics*. Berlin: Springer.

Boolos, G. (1993). *The logic of provability*. Cambridge: Cambridge University Press.

Bourget, D., & Chalmers, D. (2014). What do philosophers believe. *Philosophical Studies,170*(3), 465–500.

Brandom, R. (1994). *Making it explicit. Reasoning, representing, and discursive commitment*. Cambridge: Harvard University Press.

Burgess, J. P. (2009). Friedman and the axiomatization of Kripke’s theory of truth. Paper delivered at the Ohio State University conference in honor of the 60th birthday of Harvey Friedman.

Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID1. *The Journal of Symbolic Logic, 55,* 244–259.

Cantini, A. (1996). *Logical frameworks for truth and abstraction*. Amsterdam: Elsevier.

Carnap, R. (1931). Überwindung der Metaphysik durch logische Analyse der Sprache. *Erkenntis*, *2*(1), 219–241.

Cieśliński, C. (2007). Deflationism, conservativeness and maximality. *Journal of Philosophical Logic, 36,* 695–705.

Dummett, M. (1959). Truth. *Proceedings of the Aristotelian Society, 59,* 141–162.

Eberhard, S. (2013). Weak applicative theories, truth, and computational complexity. PhD Thesis. University of Bern.

Eberhard, S., & Strahm, T. (2012). Weak theories of truth and explicit mathematics. In U. Berger, H. Diener, P. Schuster, & M. Seisenberger (Eds.), *Logic, construction, computation* (pp. 157–184). Berlin: De Gruyter.

Eklund, M. (2001). Inconsistent languages. *Philosophy and Phenomenological research, 64*(2), 251–276.

Feferman, S. (1984). Towards useful type-free theories I. *Journal of Symbolic Logic, 49,* 75–111.
Feferman, S. (1989). Kurt Gödel : Conviction and caution. In S. G. Shanker (Ed.), *Gödel's theorems in focus* (pp. 96–115). London: Routledge.
Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic, 56,* 1–49.
Feferman, S. (1996). Gödel's program for new axioms: Why, where, how and what? In P. Hájek (Ed.), *Gödel '96*, volume 6 of *Lecture notes in logic* (pp. 3–22). Berlin: Springer.
Feferman, S. (2008). Axioms for determinateness and truth. *Review of Symbolic Logic, 1*(2), 204–217.
Feferman, S., & Strahm, T. (2000). The unfolding of non-finitist arithmetic. *Annals of Pure and Applied Logic, 104*(1–3), 75–96.
Feferman, S., & Strahm, T. (2010). Unfolding finitist arithmetic. *Review of Symbolic Logic, 3*(4), 665–689.
Field, H. (1972). Tarski's theory of truth. *Journal of Philosophy, 69,* 347–375.
Field, H. (1980). *Science without numbers: A defense of Nominalism*. Princeton: Princeton University Press.
Field, H. (2001). *Truth and the absence of fact*. Oxford: Oxford University Press.
Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.
Fodor, J. A. (1989). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge: MIT Press.
Frege, G. (1918). Der gedanke: Eine logische untersuchung. *Beitrge zur Philosophie des Deutschen Idealismus, 1,* 58–77.
Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic, 33,* 1–21.
Fujimoto, K. (2010). Relative truth definability of axiomatic theories of truth. *The Bulletin of Symbolic Logic, 16,* 305–344.
Fujimoto, K. (2012). Classes and truths in set theory. *Annals of Pure and Applied Logic, 163,* 1484–1523.
Garey, M., & Johnson, D. (2002). *Compters and intractability: A guide to the theory of NP-completeness*. New York: Freeman.
Grover, D. (1992). *A prosentential theory of truth*. Princeton: Princeton University Press.
Grover, D., Camp, J., & Belnap, N. (1975). A prosentential theory of truth. *Philosophical Studies, 27*(2), 73–125.
Gupta, A., & Belnap, N. (1993). *The revision theory of truth*. Cambridge: MIT.
Hájek, P., & Pudlak, P. (1993). *Metamathematics of first-order arithmetic*. Berlin: Springer.
Halbach, V. (1994). A system of complete and consistent truth. *Notre Dame Journal of Formal Logic, 35,* 311–327.
Halbach, V. (2010). *Axiomatic theories of truth*. New York: Cambridge University Press.
Heck, R. (2011). The strength of truth theories. (2011).
Hilbert, D. (1926). Über das unendliche. *Mathematische Annalen, 95,* 161–190.
Hindley, R., & Seldin, J. (2008). *Lambda calculus and combinators: An introduction*. Cambridge: Cambridge University Press.
Horsten, L. (1995). The semantical paradoxes, the neutrality of truth, and the neutrality of the minimalist theory of truth. In P. Cortois (Ed.), *The many problems of realism* (pp. 173–187). Tilburg: Tilburg University Press.
Horsten, L. (2011). *The Tarskian turn: Deflationism and axiomatic truth*. Cambridge: MIT Press.
Horwich, P. (1998a). *Meaning*. New York: Oxford University Press.
Horwich, P. (1998b). *Truth*. New York: Oxford University Press.
Horwich, P. (2006). The value of truth. *Nous, 40*(2), 347–360.
Kahle, R. (1999). Frege structures for partial applicative theories. *Journal of Logic and Computation, 9*(5), 683–700.
Kahle, R. (2001). Truth in applicative theories. *Studia Logica, 68*(1), 103–128.
Kahle, R. (2003). Universes over Frege structures. *Annals of Pure and Applied Logic, 119*(1–3), 191–223.

Kahle, R. (2009). The universal set—A (never fought) battle between philosophy and mathematics. In O. Pombo & Á. Nepomuceno (Eds.), *Lógica e Filosofia da Ciência*, volume 2 of Colecção Documenta (pp. 53–65). Lisboa: Centro de Filosofia das Ciências da Universidade de Lisboa.

Kahle, R. (2011). The universal set and diagonalization in Frege structures. *Review of Symbolic Logic, 4*(2), 205–218.

Kaye, R. (1991). *Models of Peano arithmetic*. Oxford: Clarendon Press.

Keefe, R. (2000). *Theories of vagueness*. Cambridge: Cambridge University Press.

Ketland, J. (1999). Deflationism and Tarski's paradise. *Mind, 108,* 69–94.

Kotlarski, H., Krajewski, S., & Lachlan, A. (1981). Construction of satisfaction classes for nonstandard models. *Canadian Mathematical Bulletin, 24,* 283–293.

Krajewski, S. (1976). Non-standard satisfaction classes. In W. Marek, M. Srebrny, & A. Zarach (Eds.), *Set theory and hierarchy theory*, Lecture notes in mathematics 537 (pp. 121–144). Berlin: Springer.

Kreisel, G. (1970). Principles of proof and ordinals implicit in given concepts. In A. Kino, J. Myhill, & R. Vesley (Eds.), *Intuitionism and proof theory* (pp. 489–503). Amsterdam: North-Holland.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72,* 690–716.

Lachlan, A. (1981). Full satisfaction classes and recursive saturation. *Canadian Mathematical Bulletin, 24,* 295–297.

Leeds, S. (1978). Theories of reference and truth. *Erkenntis, 13*(1), 111–129.

Leigh, G. (2013). A proof-theoretic account of classical principles of truth. *Annals of Pure and Applied Logic, 164,* 1009–1024.

Leigh, G., & Nicolai, C. (2013). Axiomatic truth, syntax and metatheoretic reasoning. *Review of Symbolic Logic, 6*(4), 613–636.

Leigh, G., & Rathjen, M. (2010). An ordinal analysis for theories of self-referential truth. *Archive for Mathematical Logic, 49,* 213–247.

Leigh, G., & Rathjen, M. (2012). The Friedman-Sheard programme in intuitionistic logic. *Journal of Symbolic Logic, 77,* 777–806.

Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic, 34,* 155–192.

Loewer, B. (1997). A guide to naturalizing semantics. In B. Hale & C. Wright (Eds.), *A companion to the philosophy of language* (pp. 108–126). Oxford: Blackwell Publishers.

Lynch, M. (2009). *Truth as one and many*. Oxford: Oxford University Press.

Mancosu, P. (2008a). Quine and Tarski on nominalism. In D. Zimmerman (Ed.), *Oxford studies in metaphysics* (Vol. 4). New York: Oxford University Press.

Mancosu, P. (2008b). Tarski, Neurath, and Kokoszynska on the semantic conception of truth. In D. Patterson (Ed.), *New essays on Tarski and philosophy*. New York: Oxford University Press.

Mancosu, P. (2009). Tarski's engagement with philosophy. In S. Lapointe, J. Wolenski, M. Marion, & W. Miskiewicz (Eds), *The golden age of Polish philosophy*, volume 16 of *Logic, epistemology, and the unity of science* (pp. 131–153). Amsterdam: Springer Netherlands.

Maudlin, T. (2004). *Truth and paradox*. Oxford: Oxford University Press.

McGee, V. (1992a). Maximal consistent sets of instances of Tarski's schema (T). *Journal of Philosophical Logic, 21,* 235–241.

McGee, V. (1992b). Maximal consistent sets of instances of Tarski's schema (t). *Journal of Philosophical Logic, 21,* 235–241.

McGee, V. (2006). In praise of the free lunch: Why disquotationalists should embrace compositional semantics. In T. Bolander, V. F. Hendricks, & S. A. Pedersen (Eds.), *Self-reference* (pp. 95–120). Stanford: CSLI Publications.

Moltmann, F. (2003). Nominalizing quantifiers. *Journal of Philosophical Logic, 32,* 445–481.

Moltmann, F. (2013). *Abstract objects and the semantics of natural language*. Oxford: Oxford UP.

Mulligan, K. (2010). The truth predicate vs the truth connective. On taking connectives seriously. *Dialectica, 64,* 565–584.

Nicolai, C. (forthcoming). Deflationism and the Ontology of Expressions: An Axiomatic Study. DPhil Thesis, University of Oxford.

Patterson, D. (2008). Understanding the liar. In J. C. Beall (Ed.), *Revenge of the liar: New essays on the paradox* (pp. 387–422). New York: Oxford University Press.
Patterson, D. (2009). Inconsistency theory of semantic paradox. *Philosophy and Phenomenological Research, 79*(2), 387–422.
Priest, G. (1994). The structure of the paradoxes of self-reference. *Mind, 103*(409), 25–34.
Priest, G. (2002). *Beyond the limits of thought*. Oxford: Oxford UP.
Quine, W. v. O. (1960). *Word and object*. Cambridge: MIT Press.
Quine, W. v. O. (1970). *Philosophy of logic*. Cambridge: Harvard University Press.
Quine, W. v. O. (1976). *The ways of paradox, and other essays*. Cambridge: Harvard University Press.
Quine, W. v. O. (1990). *The pursuit of truth*. Cambridge: Harvard Unviersity Press.
Ramsey, F. (1991). On truth. *Episteme, 16,* 1–16.
Rouilhan, Ph. de. (2012). In defense of logical universalism: Taking issue with Jean van Heijenoort. *Logica Universalis, 6*(3–4), 553–586.
Russell, B. (1906). On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society,* series *2*(4), 29–53.
Shapiro, S. (1983). Conservativeness and incompleteness. *Journal of Philosophy, 80*(9), 521–531.
Shapiro, S. (1998). Proof and truth: Through thick and thin. *Journal of Philosophy, 95*(10), 493–521.
Sheard, M. (1994). A guide to truth predicates in the modern era. *Journal of Symbolic Logic, 59,* 1032–1054.
Sher, G. (1991). *The bounds of logic*. Cambridge: MIT Press.
Tarski, A. (1983). The concept of truth in formalized languages. In J. Corcoran (Ed.), *Logic, semantics, metamathematics* (2nd ed., pp. 152–278). Indianapolis: Hackett.
Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic, 7,* 143–154.
Tarski, A., & Vaught, R. (1956). Arithmetical extensions of relational systems. *Compositio Mathematica, 13,* 81–102.
Tye, M. (1994). Sorites paradoxes and the semantics of vagueness. In T. E. Tomberlin (Ed.), *Philosophical perspectives 8: Logic and language* (pp. 189–206). Atascadero: Ridgeview.
Welch, P. (2003). On revision operators. *Journal of Symbolic Logic, 68,* 689–711.
Wolenski, J. (2009). The rise and development of logical semantics in Poland. In S. Lapointe, J. Wolinski, M. Marion, & W. Miskiewicz (Eds.), *The golden age of polish philosophy*, volume 16 of *Logic, epistemology, and the unity of science*. Dordrecht: Springer.
Wolenski, J., & Murawski, R. (2008). Tarski and his Polish predecessors on Truth. In D. Patterson (Ed.), *New essays on Tarski and philosophy*. Oxford: Oxford University Press.
Zardini, E. (2011). Truth without contra(di)ction. *Review of Symbolic Logic,* 4, 498–535.