All There Is

ACTA PHILOSOPHICA GOTHOBURGENSIA 31

All There Is

On the Semantics of Quantification over Absolutely Everything

Martin Filin Karlsson

UNIVERSITY OF GOTHENBURG
ACTA UNIVERSITATIS GOTHOBURGENSIS

Thesis submitted for the Degree of Doctor of Philosophy in Theoretical Philosophy
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg

ISBN 978-91-7346-949-4 (PRINT)
ISBN 978-91-7346-950-0 (DIGITAL)
ISSN 0283-2380

The publication is also available in fulltext at:
http://hdl.handle.net/2077/54485

Distribution:
ACTA UNIVERSITATIS GOTHOBURGENSIS
Box 222, 405 30 Göteborg, Sweden
acta@ub.gu.se

Typeset in Adobe Garamond Pro using XeLaTeX
Cover design by Peter Johnsen

Printed by Brandfactory Göteborg 2017

Abstract

Title:	All There Is: On the Semantics of Quantification over Absolutely Everything
Author:	Martin Filin Karlsson
Language:	English (with a summary in Swedish)
Department:	Philosophy, Linguistics and Theory of Science
Series:	Acta Philosophica Gothoburgensia 31
ISBN:	978-91-7346-949-4 (print)
ISBN:	978-91-7346-950-0 (digital)
ISSN:	0283-2380
Keywords:	absolute generality, unrestricted quantification, NFU, model-theoretic semantics

This thesis concerns the problem of providing a semantics for quantification over absolutely all there is. Chapter 2 argues against the common view that Frege understood his quantifiers in *Begriffsschrift* to range over all objects and discusses Michael Dummett's analysis of the inconsistent system of *Grundgesetze*, which generalises into his famous argument against absolute quantification from indefinite extensibility. Chapter 3 explores the possibility to adapt Tarski's first definition of truth to hold for sentences with absolute quantification. Taking the concept of logical consequence into account results in an argument for adopting a set-theory with an ill-founded membership relation as a metatheory. Chapter 4 reviews and deflates an influential argument due to Timothy Williamson against the coherence of absolute quantification. Chapter 5 discusses three important contemporary semantic theories for absolute quantification that tackle Williamson's argument in different ways. Chapter 6 challenges the widespread view that it is impossible to give a model-theoretic semantics for absolute quantification simply by providing such a semantics in NFU_p. This semantic framework provides models with the universal class as domain. I show, furthermore, that the first-order logical consequence relation stays the same in this setting, by proving the completeness theorem for first-order logic in NFU_p.

To Anneli

Acknowledgements

I would like to express my most heartfelt gratitude to my advisors, Dag Westerståhl and Christian Bennet. Dag, it has been such a privilege to write this thesis under your supervision and I am truly grateful for your meticulous reading of my texts, your firm guidance on issues of both philosophy and logic, and, of course, for suggesting the neat subject of absolutely all there is to me. Christian, your always constructive comments and warm encouragement throughout the process of writing this thesis have truly been invaluable. Our jointly written paper, *Williamson's barber*, was a pivotal moment.

I would also like to express my gratitude to Tor Sandqvist, my opponent at the final seminar. Your many insightful comments and suggestions improved this thesis considerably.

Of the many friends and colleagues that have encouraged and supported me over the years I want to mention a few in particular. Martin Kaså, my, until recently, fellow graduate student, besides everything else, I am forever in debt for the inspirational evenings we spent in the beginning of this year, working on our dissertations—you took the lead, and for me that marked the beginning of a long-awaited end. Fredrik Engström, your always generous and knowledgeable advice has put me back on track so many times. My deepest appreciation for letting so many of my ideas, and in particular those of Chapter 6, pass through your sharp mind. Rasmus Blanck, thank you so much for your careful proofreading of my manuscripts, for your patient support with all kinds of LaTeX-related stuff and for all your help and support with all other kinds of matters. Peter Johnsen, thank you very much for the cover of this book—you know what it means to me. Paul Gorbow, I am in debt for your insightful comments on the technical parts of chapter 6. Björn Haglund, you used to say that it is important to write a dissertation that is fun to read. I hope I have not let you down!

I also want to thank the participants of the logic seminar and the higher seminar in theoretical philosophy over the years.

Jan Willner, my first philosophy teacher, I hope that you enjoy to see the dissertation of an old student of yours in print.

My dear Father and my dear Brother, you've always reminded me that besides philosophy there are other important aspects of life, so now when my thesis is done, let's go watch some more football, shall we?

Astrid and Disa, having you in my life is so much joy and happiness. Anneli, my dear wife, my everything, I dedicate this book to you.

Göteborg, December 2017
Martin Filin Karlsson

Contents

1 Introduction

We seldom intend to speak about all there is. On the contrary, in our everyday conversations, quantification is almost always restricted in one way or the other; be it implicitly, by some background domain given by context, or explicitly, by some syntactic mechanism. But there are occasions when we strive to quantify over absolutely everything. For example, only an uncharitable interpreter would understand a metaphysician as quantifying over less than absolutely everything when explaining that "Everything belongs to some ontological category". Likewise, a set-theorist explaining that "nothing is a member of the empty set", does not mean to use the quantifier as restricted to some limited domain outside of which there lurk potential members that would make the set non-empty after all. Even more obviously, the Aristotelian law of identity has no bite unless it applies to everything.

But even though absolute quantification seems to be present in natural languages it is nevertheless framed with difficulties. The most challenging problem for absolute quantification stems from the classical paradoxes. Thus, Cantor's paradox of the greatest cardinal, Burali-Forti's paradox of the greatest ordinal, and Russell's paradox of classes, have all been used to argue that the very idea of absolute quantification is incoherent. Typically, this kind of argument assumes that quantification requires a domain of the things quantified over, and that the reasoning in the paradoxes shows that any such domain can always be extended to a larger domain. Michael Dummett calls such concepts *indefinitely extensible* and Russell (1907) calls the classes, or extensions, corresponding to such concepts *self-reproductive*.[1] According to this line of argument, then, there will be no such thing as a truly universal domain, and hence, nothing like absolute quantification.

[1] The notion of indefinite extensibility is recurrent in Dummett's writings; our main source here is Dummett (1991). See also Shapiro and Wright (2006) for a rewarding discussion.

A natural response to the argument from indefinite extensibility would be to accept the conclusion that, despite appearances, quantification is always limited to something less than absolutely everything. The position that quantification is always limited is often referred to as *generality-relativism*. The opposite view, i.e., that quantification is not always thus limited, is called *generality-absolutism*.

However, the position of a generality-relativist is severely problematic. Timothy Williamson shows in his thought-provoking and highly influential *Everything* (2003) that, not only is the relativist incapable of coherently articulating his position, he is also incapable of providing adequate accounts for kind generalisations and, more importantly, truth and meaning. For instance, Williamson shows that, given some very natural assumptions on context and natural languages, the relativist cannot state the truth conditions for context-sensitive universally quantified sentences in a context-sensitive metalanguage. What the relativist wants to say is that,

(∗) for any context C, and any sentence of the form $\forall x\varphi$, $\forall x\varphi$ is true in C if and only if, every member of the domain of C satisfies φ in C.

But since quantification in the metalanguage is context-sensitive, the context in which (∗) is uttered, CT say, provides a domain. Thus, as Williamson points out, for some particular context C instantiating (∗), the resulting condition would be that $\forall x\varphi$ is true in C if and only if, every member of the intersection of the domain of C and the domain of CT satisfies φ in C. Hence, in order to get the right truth conditions for each context C, the domain of CT needs to contain all members of the domain of every C in which $\forall x\varphi$ may be uttered. But this looks dangerously close to asking for a context with a domain of everything there is or, alternatively, that there is something not within any of the domains of the possible contexts for $\forall x\varphi$. But that there would be some object outside of each possible context for $\forall x\varphi$, if $\forall x\varphi$ stands for a natural language sentence, seems highly implausible, and is in any case not an option for a generality-relativist.

This and a number of other pertinent arguments makes Williamson claim that generality-relativism leads to meta-linguistic pessimism, that is, "it endangers the possibility of a reflective understanding of our own thought and language, even from the standpoint of a meta-language."[2] The task for the generality-absolutist, then, is to show that, given his understanding of the quantifiers, coherent meta-linguistic reflection is in fact possible.

Interestingly, Williamson claims that model-theoretic semantics is problematic in a similar way:

> Speaking in the metalanguage of first-order model theory, one says: every model has a set for its domain; since no set contains everything, no model has everything in its domain; but each thing belongs to some sets (such as its own singleton) and therefore to the domain of some model or other. Consequently, no model has every model in its domain. Thus a formalization of the meta-theory in a first-order language has no intended model, in the standard sense. (Williamson, 2003, p. 446)

Standard model-theoretic semantics is inadequate for absolute quantification according to Williamson, not only because of the lack of a universal set for the domains of the models, but also because of the lack of an intended model for the semantic theory itself.

In this thesis I will argue that we don't have to give up the idea of a first-order model-theoretic semantics for absolute quantification. That is, I argue that the absolutist may indeed construct a model-theoretic semantics to meet Williamson's challenge. A first-order formulation of such a semantics will contain 'model', 'assignment' and 'satisfaction' among its predicates and relations. A model, $\mathcal{M}^{\Pi}$, for such a language will, like any model for a first-order language, have a domain of quantification, M^{Π}, and a function I^{Π} interpreting the predicates and relation symbols. I define $\mathcal{M}^{\Pi}$ in the set theory $\mathrm{NFU_p}$ that results from Quine's NF if we add urelements and, for convenience, a primitive pairing operator, in Chapter 6.[3] The

[2] Williamson (2003, p. 452). But see also Studd (2017) for a defence of generality-relativism with regard to semantic pessimism.

[3] 'NF' abbreviates the title of Quine's *New Foundations for Mathematical Logic* (1937). The theory that results from adding urelements, NFU, was first suggested in Jensen (1968-69), wherein the consistency of NFU, NFU with infinity and choice, is proven (relative

models in M^{Π} that have the universal set as their domain of quantification will serve as interpretations of absolute quantification. Furthermore, the resulting semantics is shown to be complete with regard to standard proof-systems. Thus the concept of *first-order consequence* in the new sense will be co-extensional with the concept of *derivability*, which, in turn, implies that it is co-extensional with the standard model-theoretic concept of consequence as well. Thus, despite using a non-standard set theory, the resulting model-theoretic semantics will be standard in this important respect.

But *can* does not imply *ought* and the construction of a model-theoretic semantics in $\mathrm{NFU_p}$ needs motivation. Chapters 2–5 are meant to provide, in various ways, a motivation of sorts, partly by commenting on discussions and objections in the literature. Below I summarise the main theme of each of these chapters.

A good starting point for a discussion on absolute quantification is the works of Gottlob Frege. One reason is that he adopted a logical system in his *Grundgesetze der Arithmetik* (1893,1903) that allowed the derivation of Russell's paradox. Hence, he inadvertently provided one of the most influential arguments against absolute quantification. Another reason is that he, according to the common view, employed, or intended to employ, absolute quantification by taking his first-level quantifiers to range over absolutely all objects.

Frege was clear about the syntax and, to some extent, the semantics of quantification already in his first book on logic, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* (1879). In the first part of Chapter 2 I challenge the common view that Frege took his quantifiers as ranging over absolutely everything in that book, by arguing that they are best understood as substitutional. That is, rather than saying that the sentence $\forall x\varphi$ is true if φ is true for all *values* of x, I claim that the quantifiers in *Begriffsschrift* render it true if φ is true for all legitimate substitution instances of x. Hence, since we only quantify over named objects—if any-

to ZFC). Type levelled ordered pairs are defined for NF in Quine (1945) but that definition assumes infinity and is thus not suitable in NFU. A primitive type levelled pairing operator was added to NFU in Feferman (1974). The resulting theory is referred to as $\mathrm{NFU_p}$.

thing at all—when applying substitutional quantification, the only further assumption needed in order to refute absolute quantification (objectually understood) is that some objects have no names.

In the second part of Chapter 2 I discuss Dummett's influential argument from *indefinite extensibility* in the historical context of *Grundgesetze*. Richard Cartwright's *Speaking of Everything* (1994) is an interesting response to Dummett's argument. Cartwright argues that it is misguided to assume, as Dummett needs to do, that there ought to be a completed collection of the things we quantify over in addition to the things themselves. Cartwright's response is interesting in its own right, but it also gives us reason to discuss the relation between model-theoretic semantics and the ontological commitments in the object language. One worry is that, since we quantify over domains in the metalanguage, an object language with absolute quantification will inherit commitments to domains from the metalanguage. We close Chapter 2 by sorting out this question.

To provide a model-theoretic semantics one needs to define the relation, $\mathcal{M} \models \varphi$, of a sentence φ being true in a model $\mathcal{M}$. Alfred Tarski is rightly acknowledged for the now standard definition of truth in a model, but he did not give that definition, as is sometimes claimed, in his *The Concept of Truth in Formalized Languages* (1935).[4] In that work Tarski defines plain truth for interpreted formalised languages. Interestingly, Tarski imposes no explicit restriction on the quantifiers in his definition and it is therefore tempting to try to adapt his method to languages with absolute quantification. We discuss this possibility in the first part of Chapter 3 and find that the main obstacle is Tarski's use of Husserl's semantic categories in the metalanguage. Since the variables in the metalanguage for variable assignments and the variables in the object language necessarily belong to different categories, it seems in principle impossible for the object language quantifiers to range over the variable assignments. Thus, from the perspective of the metalanguage, there is something over which the object language quantifiers does not range, and hence they fall short of being absolute.

The second part of Chapter 3 provides two alternative ways of circum-

[4] The first printed definition of truth in a model seems to be that in Tarski and Vaught (1957). See Hodges (1985/6) for a discussion.

venting the problem raised by using Husserl's semantic categories in the context of absolute quantification. The first alternative is to argue that Husserl's semantic categories bring no new ontology. One can do that in the same way as higher-order languages are sometimes said to bring new ideology (or expressive power) rather than new ontological commitments in addition to those following from the first-order quantifiers. This comes close to using semantic theories based on some ontologically innocent type theory. We discuss two such theories in Chapter 5. The second alternative consists in trading Husserl's categories for ZF.

Much of the interest in model-theoretic semantics lies in its ability to provide adequate definitions of logical relations between sentences. Of fundamental interest is the relation of logical consequence. In the third part of Chapter 3 we show that the semantics that results from trading Husserl's categories for ZF, although it provides a truth definition, fails with regard to the definition of logical consequence. One interesting reason for this is the axiom of foundation in ZF, which makes $\in$ well-founded. This suggests that we should use, as we do in Chapter 6, a set theory where $\in$ is not well-founded as our metatheory.

Dummett's argument from indefinite extensibility against the coherence of absolute quantification is set-theoretic in spirit. Accordingly it is highly dependent on one's adopted set theory. However, Williamson (2003) gives an argument in the same spirit that makes no assumptions on sets or classes. The argument is presented as a reductio of the assumption of absolute quantification. In addition to the assumption of absolute quantification the argument uses two further premisses. The first premise is that there is an interpretation that interprets a predicate of the object language in accordance with any possible semantic value suitable for that predicate; in particular this holds for the semantic values of the predicates in the (interpreted) metalanguage. The second premise is that a definition of a particular predicate in the metalanguage is legitimate. Even though it involves no assumptions on sets or classes it has been analysed, notably in Glanzberg (2004) and Parsons (2006), in terms of indefinite extensibility.

Williamson's argument plays a major role in the contemporary discussion of absolute quantification and I devote Chapter 4 to it. Having presented the argument I review Glanzberg's and Parsons' analyses in terms of indefinite extensibility. I also give an analysis that is closer to Dummett's use of indefinite extensibility. Finally I show that the argument is best understood as a reductio of the definition of the purported predicate in the metalanguage rather than the assumption of absolute quantification. As long as we don't impose some principle making that faulty definition legitimate the project of giving a model-theoretic semantics for absolute quantification is unhampered by Williamson's argument. This chapter relies on joint work with Christian Bennet.

The theories that have been proposed as a response to Williamson's challenge of constructing a semantics for absolute quantification have all taken Williamson's argument seriously. They also adhere to the requirement that a semantics should be *strictly adequate*, roughly in the sense that for any possible semantic value a predicate may have, there ought to be some interpretation that interprets it accordingly, or *general* in the sense that it should be applicable to any legitimate first-order language. In Chapter 5 I critically review three theories aiming at meeting these requirements in different ways. Two of the theories are type-theoretical. Thus, both Williamson (2003) and Rayo (2006) suggest that we should use higher-order resources to set up a semantic theory, while they differ in their interpretations of the higher-order quantifiers. Williamson suggests that we should take our higher-order quantifiers as ranging over concepts and Rayo suggests that they can be interpreted in a higher-order plural language. Both Williamson and Rayo resist any claim that higher-order quantification entails commitment to entities in addition to the entities that the first-order quantifiers range over.

Williamson's and Rayo's accounts require an infinite hierarchy of languages of different orders and some well-known problems of stating truths about such hierarchies from somewhere within the hierarchy become relevant. The inability to express certain truths makes it look like these theories are at the brink of violating the idea of strict adequacy. This shows that, although it is possible to show that the hierarchy provides a strictly adequate semantics for each level in the hierarchy, there is no strictly adequate semantics for the language consisting of all levels. Moreover, I argue, the notion of strict adequacy itself cannot be expressed from within the hierarchies. It is therefore doubtful that the idea of strict adequacy can be used to motivate the higher-order approach. Furthermore, I present a dilemma that the type-theorist faces if we are allowed to form predicates of the metalanguage in any way we want. For then we may construct a contradiction that shows, either that the higher-order account is contradictory, or that some predicate is illegitimate, which would make a similar response to Williamson's argument possible and hence neutralise an important argument for turning to higher-order resources in the first place.

Due to the lack of a universal set and the set-theoretic paradoxes, standard set theory, ZFC, constitutes a poor framework for a semantics adequate for absolute quantification. A natural response is thus to give up ZFC as a metatheory. However, Linnebo (2006) takes a stand against such a conclusion. Rather than giving up ZFC he suggests that it should be supplemented with a theory of properties. According to Linnebo, the resulting theory must be strong enough to construct an adequate semantics and the theory of properties must avoid Williamson's argument in a natural, non ad hoc way.

Inspired by Linnebo's first-order approach I proceed in Chapter 6, after a brief discussion of the most common objections, to construct the model-theoretic semantics mentioned above.

2 Frege: General Statements and Quantification

In the last section of Rayo and Williamson (2003) the following historical note is made:

> The formal system which Frege set forth in the *Begriffsschrift* was meant to be a universal language; it was intended as a vehicle for formalising all deductive reasoning. Accordingly, Frege took the first-order variables of his system to range over *all* individuals. So much is beyond dispute. (Rayo and Williamson, 2003, p. 354, italics in the original)

This view on general judgements in the *Begriffsschrift* is not uncommon, but, as stated above, it carries an ambiguity. '*Begriffsschrift*' is sometimes used to abbreviate the title of Frege's first book on logic, which is how we will use the word, and sometimes to refer to the formal language of that book. Frege used that language, with important amendments, in later texts, e.g., *Grundgesetze*.[1] Though the symbols and formation rules used in *Begriffsschrift* are fully incorporated in *Grundgesetze*, they are given a radically different semantics in that later work. Indeed, the differences are so profound that one can hardly speak of *one* formal system. Hence it is imprecise to speak of *the Begriffsschrift*, as if there is only one; that is, unless one intends to refer to Frege's first book on logic.

It follows that there are at least two ways of understanding the second sentence of the quote, depending on whether it concerns the first-order variables in *Begriffsschrift* or in *Grundgesetze*. In this chapter we investigate both alternatives.

In Section 2.1 we present an interpretation of *Begriffsschrift* according to which the first-order variables, or equivalents of such variables, may *not* be

[1] When referring to Frege's texts I use *Begriffsschrift* to refer to the translation Frege (1879), and *Grundgesetze* to refer to the edited translation Frege (1893,1903).

understood to range over absolutely all individuals. This follows from an argument showing that quantification in *Begriffsschrift*, properly understood, is substitutional (as opposed to objectual). Thus, only named objects are quantified over. Given the plausible assumption that not all individuals have names, the non-absoluteness of the quantifiers follows.

This is a small historical observation in the over-all discussion of the question of the possibility of an absolutely general enquiry. It gives a negative answer, for a particular formal system, to a specific interpretation of what Rayo and Uzquiano calls the availability question: "Could an all-inclusive domain be available to us as a domain of enquiry?" (Rayo and Uzquiano, 2006, p. 2) If Frege wanted his quantifiers to range over absolutely everything, a natural and simple solution would have been to keep the formal part of *Begriffsschrift* and interpret quantification objectually instead. Actually, this is roughly what Frege does in *Grundgesetze*, although for other reasons. A large part of the formal system is kept intact while some new notation is introduced to match his new semantics which had become much more involved.

However, due to the derivability of Russell's paradox in *Grundgesetze*, the assumption of variables ranging unrestrictedly becomes non-trivial.[2] The paradox shows that the system of *Grundgesetze* is inconsistent. Michael Dummett has argued that Frege's mistake lay in the failure of acknowledging the existence of indefinitely extensible concepts.[3] Roughly, a concept is said to be indefinitely extensible if the hypothesis that it has a determinate extension gives rise to entities that, although falling under the concept, cannot belong to the extension. As an example, Russell's paradox may be used to show that the notion of *set* is indefinitely extensible. Assume thus that the concept *set* has a definite extension E. Then, among the Es, there will be sets that are not members of themselves, and, using comprehension, we may form the set of all those sets that are not members of themselves. Russell's paradox then shows that this set cannot be among the Es. Consequently, E did not consist of all sets.

[2] Note that the formal system of *Begriffsschrift* is consistent. For a proof, see Russinoff (1987).

[3] Dummett argues along such lines in a number of places, but we mainly confine ourselves to Dummett (1991).

This actualises the metaphysical counterpart of the availability question, "Is there an all-inclusive domain of discourse?" (Rayo and Uzquiano, 2006, p. 2). These two questions form the core of the modern discussion of absolute generality. Thus, the problems identified in the modern discussion of the possibility of an absolutely general inquiry were present at the dawn of modern quantificational logic.

In Section 2.2 we look at the concept of indefinite extensibility as applied to *Grundgesetze* and in what way it may be thought of as a problem for absolute quantification. We also consider an argument by Cartwright (1994) according to which this problem is not insurmountable. Furthermore, we discuss what consequences indefinite extensibility and Cartwright's argument might have for a putative model-theoretic semantics for absolute quantification.

2.1 General statements in *Begriffsschrift*

Begriffsschrift consists of three parts. In the first part a general presentation of the notational system is given; in the second part, fifty-nine propositions are derived from nine axioms by means of modus ponens and substitution; in the third part, finally, some general propositions about sequences are derived from four definitions and the propositions derived in the second part.

Our main concern here is universal quantification and the question if Frege took his quantifiers to range over absolutely everything in *Begriffsschrift*. We start in Section 2.1.1, by rehearsing some basic notions and distinctions. In Section 2.1.2 we briefly discuss the much debated distinction between function and argument in *Begriffsschrift*. As it turns out, the argument given in Section 2.1.3 is intertwined with this thorny debate. What we show in Section 2.1.3 is, basically, that the way Frege introduces the identity sign in *Begriffsschrift* gives us a strong reason for adopting a substitutional reading of the quantifiers. In Section 2.1.4, we discuss what the substitutional reading implies with regard to absolute quantification. We find that it is implausible to regard quantification in *Begriffsschrift* as absolute.

2.1.1 Preliminaries

A reader of *Begriffsschrift* who is only familiar with Frege's later texts may find its semantics rather crude and unanalysed. For instance, the distinction between an expression having a *Bedeutung* and expressing a *Sinn*, introduced in *Über Sinn und Bedeutung* (1892), is lacking. Instead, in retrospect, these semantic categories are merged into the one single category of *content* ('Inhalt').

The conditional and negation are explicitly designated in the formal language, call it $\mathscr{L}_{Bs}$, of *Begriffsschrift* and are understood to operate on contents. The affirmation of A's content standing in the conditional relation to B's content is explained as (i) the affirmation of the content of A and the affirmation of the content of B, or (ii) the affirmation of the content of A and the denial of the content of B, or (iii) the denial of the content of A and the denial of the content of B.[4] Hence, rather than being truth-functional, the sentential calculus in *Begriffsschrift*, that is, the calculus involving only negation and the conditional, is content-functional. Furthermore, the sentential calculus is compositional: given the contents of the parts of an expression of a complex sentence in $\mathscr{L}_{Bs}$, the content of the whole expression is a function of those parts and their mode of composition.

Frege distinguishes signs for logical constants from letters:

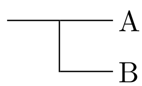
> I [...] divide all signs that I use into *those by which we may understand different objects* and *those that have a completely determinate meaning.* The former are *letters* and will serve chiefly to express *generality*. But no matter how indeterminate the meaning of a letter, we must insist that throughout a given context the letter *retain* the meaning once given to it. (Frege, 1879, p. 10, italics in the original)

It is unfortunate that Frege does not take the opportunity to list, or at least exemplify, the two kinds of symbols he considers in this passage.

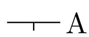
Frege employs a horizontal stroke to indicate that we *use* an expression A of $\mathscr{L}_{Bs}$, instead of using inverted commas in order to *mention* it. That is,

[4] Frege also requires that the contents of A and B are *judgeable*, but we may ignore this slight complication here.

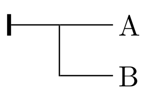
the horizontal stroke takes us from the expression to its content, very much as one may say that inverted commas take us from a symbol to its name. Thus '—— A' marks the content expressed by A. The conditional relation between two contents expressed by A and B is then readily expressed:

(1) 

In (1) the vertical stroke marks that the content of A is implied by the content of B. The leftmost horizontal stroke is the content stroke for the combination of signs to the right of it. Negation of the content of A is expressed by a small vertical stroke dividing A's content stroke:

(2) 

The affirmation of a content is marked by a vertical stroke added to the leftmost content stroke. Thus

(3) 

expresses the *affirmation* of the content of the conditional.

The use of signs for the conditional and negation is thus unambiguously explained in *Begriffsschrift*. Unfortunately, this is not true for Frege's use of letters like 'A' and 'B'. We are told in a footnote that the capital Greek letters employed ('A' is a capital alpha and 'B' a capital beta) are "abbreviations" and that we "should attach an appropriate meaning when I do not expressly give them a definition" (Frege, 1879, p. 11 n.). However, in the course of reading *Begriffsschrift* it becomes reasonably clear that capital Greek letters are used as schematic letters for expressions in $\mathscr{L}_{Bs}$.[5]

[5]Not everyone agrees. Baker and Hacker (1984, p. 171) seem to understand the Greek capitals as denoting objects and concepts, treating them as if they belonged to $\mathscr{L}_{Bs}$. It is true that capital Greek letters appear in part two and three of *Begriffsschrift*, i.e. not only in the part that explains the symbolism, but then only as schematic letters in tables for substitutions. It is natural to think of them as merely schematic also in the first part of *Begriffsschrift*.

That the meanings of Greek letters have to be "appropriate" means that whatever combination of signs in $\mathscr{L}_{Bs}$ we understand them to stand for, the combination of signs to the right of the judgement sign '$\vdash\!\!\!-\!\!-$' must have a content that is capable of becoming a judgement.

An important distinction in *Begriffsschrift* (p. 3) is that between *contents* and *conceptual contents*. Consider

(4) If John loves Mary, then John is happy.

and

(5) If Mary is loved by John, then John is happy.

Frege would say that (4) and (5) have different contents, even though this difference is of no *logical* significance. That part of the content playing a role in logical inferences Frege calls the *conceptual content* and thus he may say that (4) and (5) have the same conceptual content. A conceptual content expressed by two different natural language expressions may thus be formalised by the same expression in $\mathscr{L}_{Bs}$.

2.1.2 The function-argument distinction

Consider (4) again. If 'Mary' is viewed as replaceable, then the expression splits up into a replaceable part, 'Mary', and a constant part, 'If John loves (Mary), then John is happy'. The constant part Frege calls a *function* and the replaceable part he calls an *argument* (to the function):

> *If in an expression, whose content need not be capable of becoming a judgement, a simple or compound sign has one or more occurrences and if we regard that sign as replaceable in all or some of these occurrences by something else (but everywhere by the same thing), then we call the part that remains invariant in the expression a function, and the replaceable part the argument of the function.* (Frege, 1879, p. 22, italics in the original)

Symbolically, in the first part of *Begriffsschrift*, an arbitrary function of one argument is written $\Phi(\mathrm{A})$, and a function of two arguments as $\Psi(\mathrm{A}, \mathrm{B})$.

Now, having a natural language statement n, expressing a content, the conceptual content of which we may denote by '$C(n)$', it is possible to form a function in the above sense by considering a part of n as being replaceable. Thus, if a (one-place) function is formed from n, n is viewed as consisting of one constant part (the function) and one replaceable part (the argument), and it is only natural to assume that $C(n)$ splits up into two parts in a similar fashion. Indeed, this is more or less Frege's standpoint in his later texts were he explicitly distinguishes functions from *objects* in the contents of statements.[6] However, such a distinction among the conceptual contents is not explicitly endorsed in *Begriffsschrift*.

But consider now the following quote in which 'Cato killed Cato' is analysed:

> If we here think of "Cato" as replaceable at its first occurrence, "to kill Cato" is the function; if we think of "Cato" as replaceable at its second occurrence, "to be killed by Cato" is the function; if, finally, we think of "Cato" as replaceable at both occurrences, "to kill oneself" is the function. (Frege, 1879, p. 22)

In this quote Frege may be interpreted as speaking of parts of C(Cato killed Cato). For instance, 'to be killed by Cato' is not literally a part of 'Cato killed Cato', and hence, it does not result from the decomposition of the sentence into a function part and an argument part in accordance with the procedure described above. Instead it seems plausible to assume that it is the conceptual content of 'Cato killed Cato' that is decomposed in the quote, and that these parts are denoted by 'to be killed by Cato' and 'Cato'.

Examples like this have been taken to show that the function-argument distinction is in fact not unambiguously introduced in *Begriffsschrift*. Officially it is introduced as applicable at a syntactic level, but Frege sometimes speaks as if it is also applicable at a semantic level.

It should be said that the 'Cato killed Cato' example is not the only argument for a non-syntactic reading of the distinction. Notably Baker (2001) and Baker and Hacker (1984, 2003) provide arguments relying on close

[6] See Frege (1891).

readings of Frege's use of quotation marks in *Begriffsschrift* and elsewhere, as well as Frege's retrospective comments on the distinction. While Baker and Hacker defend a strictly non-syntactic understanding of the distinction, others disagree. Dummett (1984), for instance, argues in his detailed review of Baker and Hacker (1984) that Frege saw the distinction applicable at both a syntactic and a semantic level (p. 380), and that it is far from certain that Frege was clear about the nature of functions at the time of writing *Begriffsschrift* (p. 381).

Surprisingly, as we shall see next, a rather straightforward reading of Frege's treatment of identity and quantification sheds light on this debate. In fact, it provides a strong argument for the syntactic reading.

2.1.3 Quantification and identity

Consider the following passage of *Begriffsschrift* where Frege introduces the notation for quantification.

> In the expression of a judgement we can always regard the combinations of signs to the right of ⊢── as a function of one of the signs occurring in it. *If we replace this argument by a German letter and if in the content stroke we introduce a concavity with this German letter in it, as in*
>
> $$\vdash\!\overset{\mathfrak{a}}{\smile}\!-\ \Phi(\mathfrak{a})$$
>
> *this stands for the judgement that, whatever we may take for its argument, the function is a fact.*[7] (Frege, 1879, p. 24, italics in the original)

There are two content strokes involved in the notation for quantification. The one to the left of the concavity Frege explains to be the content stroke "for the circumstance that, whatever we may put in place of $\mathfrak{a}$, $\Phi(\mathfrak{a})$ holds" and "the horizontal stroke to the right of the concavity is the content stroke

[7] The notion of a function being a fact does not really make sense here. Either it is the value of the function that is a fact, or the content of that value, depending one's understanding of functions.

of $\Phi(\mathfrak{a})$, and here we must imagine that something definite has been substituted for $\mathfrak{a}$." (Ibid., p. 24)

Given the first sentence of the quote, where Frege explicitly talks about the "*expression* of a judgement" and "combination of *signs*," it seems quite clear that Frege gives an account of substitutional quantification. That is, anachronistically put,

$$\vdash\!\overset{\mathfrak{a}}{\smile}\!-\Phi(\mathfrak{a})$$

if and only if

$$\vdash\!\!\!\longrightarrow \Phi(\mathfrak{a}) \text{ for all substitution instances of } \mathfrak{a}.$$

Despite such textual evidence Dummett defends the view that Frege actually meant to give an account for objectual quantification:

> The much more loosely expressed stipulation in *Begriffsschrift*, §11, concerning the quantifier reads:
>
> > $\vdash\!\overset{\mathfrak{a}}{\smile}\!-\Phi(\mathfrak{a})$ signifies (*bedeutet*) the judgement that the function is a fact whatever we take as its argument.
>
> Fairly clearly, this too, is intended to express an objectual interpretation of the first-order quantifier, an interpretation that Frege appears to have put on it throughout his career. (Dummett, 1991, p. 206)

A similar view seems to be embraced by Michael Beaney who explains the notation for first-order quantification in an appendix to *The Frege Reader*:

> This is understood as representing the judgement that 'the function [Φ] yields a fact whatever is taken as its argument', i.e. that everything has the property Φ (for all $x\,Fx$ – '$\forall x Fx$' as it would be formalized in modern notation). (Beaney, 1997, p. 379)

The understanding of quantification in *Begriffsschrift* also depends on the function-argument distinction. If that distinction only applies at the level of syntax there seems to be no alternatives save for a substitutional reading of the quantifiers. On the other hand, if one thinks that the distinction does not apply on a syntactic level, or at least not *only* on a syntactic level, then the objectual reading of the quantifiers makes perfect sense.[8] Thus, quantification seems to be ambiguous in the very same way as the function-argument distinction.

We now turn to Frege's treatment of identity in *Begriffsschrift*. When introducing identity Frege speaks of it as being of another kind than the conditional and negation:

> Identity of content differs from conditionality and negation in that it applies to names and not contents. Whereas in other contexts signs are merely representatives of their content, so that every combination into which they enter expresses only a relation between their respective contents, they suddenly display their own selves when they are combined by means of the sign for identity of content. (Frege, 1879, p. 20)

The introduction of the identity sign into the formula language is motivated by the need to take care of the informativeness of sentences of the kind 'Scott is the author of Waverley', but also for enabling stipulated abbreviations (i.e. definitions) in the formula language. If identity were a relation between contents, there would be no difference between 'Scott is the author of Waverley' and 'Scott is Scott'—two judgements that arguably differ as to their content. Hence, the sign is introduced and explained:

> Now let
>
> $$\vdash A \equiv B$$
>
> mean that *the sign* A *and the sign* B *have the same conceptual content, so that we can everywhere put* B *for* A *and conversely.* (Frege, 1879, p. 21, italics in the original)

There is no doubt that Frege puts the relation of identity at the level of syntax. This is clear, both from the actual words introducing it, and from the

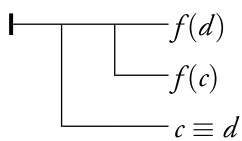
[8] Accordingly, Baker and Hacker (1984, p. 181), who favour a non-syntactic understanding of the distinction, may embrace an objectual reading.

fact that it simply wouldn't do the job if it was a relation between contents.[9]

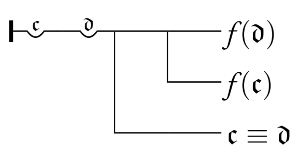
Although identity is unambiguously introduced, it is sometimes thought to raise internal problems in the formal system of *Begriffsschrift*. Thus, in the introduction to his translation of *Grundgesetze*, Montgomery Furth comments on the use of the identity sign in *Begriffsschrift*:

> It has the merit of accounting for the interest of true "$A = B$" as against the uninformativeness of "$A = A$". But the price is exorbitantly high, for the device renders it practically impossible to integrate the theory of identity into the formalised object-language itself; e.g. to state generally such a law as that if $F(a)$ and $a = b$ then $F(b)$. (p. xix, Furth's introduction to Frege (1893,1903))

In $\mathscr{L}_{Bs}$ such a general law would be expressed by

(6) 

Now, (6) is actually axiom (52) in the deductive system in part two of *Begriffsschrift*. The use of Latin letters in (6) is explained by a convention that they are universally quantified with the whole judgement as their scope (p. 21). Thus, this axiom may also be written:

(7) 

Furth's worry is that this does not express a general law because the antecedent with the identity sign seems to restrict the content of the judgement to signs. But, of course, this problem appears only if we read the quantifiers

[9] In later texts, Frege would say that in sentences such as 'Scott is the author of Waverley', 'Scott' and 'the author of Waverley' share their *Bedeutung* but have different *Sinn*. But, as we said, such a distinction is not present in *Begriffsschrift*.

objectually.

Indeed, to substantiate Furth's problem, one needs more than objectual quantification in *Begriffsschrift*. First, we must identify the parts in the conceptual contents that correspond to the arguments in the syntactic function-argument distinction. For simplicity, call such alleged parts *objects*.[10] Second, assuming that conceptual contents may have argument places, we have to acknowledge that such argument places, in case they appear in the conceptual content of an identity, must be filled by signs. That is, we need to recognise signs among the objects. Then we could say that (7) is tacitly restricted to that subcategory of objects which consists of the signs that may appropriately flank the sign for identity. That is to say, we are then in a position to claim that Frege failed, due to the syntactic character of the identity relation, to properly express such general laws as $\forall a \forall b(a = b \to (F(a) \to F(b)))$.

However, under the substitutional interpretation of quantification none of these problems appear. Rather than perceiving the relation of identity as giving rise to internal problems of the kind Furth suggests, we may use it to argue for a substitutional reading of quantification: Frege succeeds in stating axiom (52) with its intended meaning *precisely because quantification is substitutional.*

The substitutional reading of quantification implies that the function-argument distinction applies primarily at the level of syntax. But then, in order to avoid a reductio-argument, we need to explain the 'Cato-killed-Cato' example. The problem is that, taking the first occurrence of 'Cato' to be the argument, Frege explains the function to be 'to be killed by Cato', an expression which is not literally a part of the decomposed expression, i.e. of 'Cato killed Cato'. Thus, one may argue, this example shows that the function-argument distinction applies at a non-syntactic level. However, this argument is rather weak. One need only acknowledge that two different natural language expressions may have the same formal rendering in $\mathscr{L}_{Bs}$.[11] Thus, when Frege speaks of expressions, in the course of stating the process of decomposing an expression into function and arguments,

[10] This is consonant with the terminology of Baker and Hacker (1984, ch. 7).

[11] A similar point is made in Baker and Hacker (2003, p. 277).

he speaks of expressions of $\mathscr{L}_{Bs}$ rather than natural language expressions. Accordingly, in the 'Cato-killed-Cato' example we may understand 'to be killed by Cato' as 'the formalisation into $\mathscr{L}_{Bs}$ of "to be killed by Cato" '.

There is no telling what Frege's intention was regarding the quantifiers in *Begriffsschrift*. The passages introducing quantification do not give unambiguous support for either the substitutional or the objectual reading. There are other arguments for the objectual reading, but they all rely on close readings of passages in the first part of *Begriffsschrift*, or comparative readings of contemporary sources, or retrospective comments in later sources. In contrast, the above argument from the accuracy of axiom (52) is simple and straightforward, assuming nothing that isn't explicitly and officially stated in *Begriffsschrift*, and it embraces only the quite harmless presumption that Frege intended to say what he actually says in axiom (52). This is a strong argument for the view that quantification in *Begriffsschrift* is, in fact, substitutional.

2.1.4 Absolute generality in the *Begriffsschrift*.

Let us recall a standard account of substitutional quantification.[12] Assume that $\mathscr{L}$ is a first-order language and let an *interpretation* I be a mapping of atomic $\mathscr{L}$-sentences onto $\{T, F\}$. Define an *I-valuation*, v_I, by recursion on the complexity of formulas in the following way:

1. If ψ is an atomic sentence, then $v_I(\psi) = I(\psi)$, and
2. if ψ is $\neg\phi$, then $v_I(\psi) = T$ if and only if $v_I(\phi) = F$, and
3. if ψ is $\phi \vee \chi$, then $v_I(\psi) = T$ if and only if $v_I(\phi) = T$ or $v_I(\chi) = T$, and
4. if ψ is $\forall x\phi$, then $v_I(\psi) = T$ if and only if $v_I(\phi(n)) = T$ for all $n \in C$, where $\phi(n)$ is the result of substituting n for all free occurrences of x and C is some denumerable class of suitable $\mathscr{L}$-terms.

Though it is quite possible to understand substitutional quantification as ontologically committing by requiring that the items in C refer to objects

[12] The account given here is essentially the same as in Dunn and Belnap (1968).

of some sort, one of the main attractions with this kind of quantification is that such an understanding is not forced upon us. For instance, if the sentence '$\exists x(x$ defeated the Hydra)' is true under an interpretation I, this is not so because there is someone who defeated the Hydra, but because the sentence has at least one instance that v_I maps to T.

The ontological innocence of substitutional quantification has been considered an advantage since it allows for a semantics without requiring a specifiable extra-lingual domain of quantification. A truth definition for $\mathscr{L}$ requires nothing but the syntax of $\mathscr{L}$ and a mapping of atomic sentences to truth-values. Here is how Ruth Barcan Marcus defends this kind of quantification:

> The impetus for the initial proposal [of substitutional quantification] was not, as sometimes suggested, grounded in finding a way of quantifying into and out of modal contexts. [...] It is rather the much more general observation that there is a genuine question about the appropriateness or even the meaningfulness of supposing that there is a clear connection between the standard interpretation of the quantifiers and any paraphrase into and out of ordinary and philosophical discourse. The standard semantics *demands* a clearly specifiable domain over which the variables range and which are its values. [...] Then what, if we are dealing with ordinary or philosophical discourse, is the clearly specifiable domain over which the variables range? (Marcus, 1972, p. 244)

The point is thus that since we cannot always specify a domain of quantification, we shouldn't adopt a semantics requiring such a domain. Instead we should adopt a semantics free from ontological commitments, e.g., a semantics interpreting the quantifiers substitutionally.

Though it is doubtful that Frege ever thought along these lines when defining quantification in *Begriffsschrift*, it is nevertheless consistent with his exposition of the formal system expressed in $\mathscr{L}_{Bs}$. An anachronistic conclusion would then be that, contrary to the common view that quantification in *Begriffsschrift* is over absolutely everything, rather, it is over nothing.

Another question regarding substitutional quantification is if we can always decide whether we, in colloquial language, use substitutional or objectual quantification.[13] Surely, if we are able to conclude that $\neg\forall x\phi(x)$ is true even though $\phi(n)$ holds for all $n \in C$, we can safely say that we use objectual quantification. But if we find ourselves in a situation where there is a true instance $\psi(n)$ for any true existential sentence $\exists x\psi$, we will no longer be able to separate the unnamed objects from the named ones. In particular, given that C contains a witness for each true existential sentence, there is no such formula as $\phi(x)$ above; since $\neg\forall x\phi(x)$ entails $\exists x\neg\phi(x)$, there is an instance $\neg\phi(n)$ contradicting the assumption that $\phi(n)$ holds for all $n \in C$. Accordingly, by merely knowing the truth-values of the sentences considered, there would be no way of distinguishing substitutional form objectual quantification.

If we cannot tell if we use substitutional or objectual quantification, the range of our quantifiers seems to be indeterminate. There is simply no way to tell if we quantify, substitutionally or objectually, solely over named objects, or if there are also unnamed objects, inseparable from the named ones, which are anyway within the range of our quantifiers. Clearly, in the latter case, quantification ought to be objectual, even though we are not in a position to know that.[14]

The problem in *Begriffsschrift*, however, is not that we cannot tell the two types of quantification apart. Hence there is no risk of accidentally quantifying over unnamed objects. Furthermore, if we consider the expressions substituted for variables as referring, there is still an indeterminacy concerning the range of quantification in *Begriffsschrift* since the expressions allowed for substitutions are not unambiguously delineated. Despite this indeterminacy, the plausible assumption that quantification is substitu-

[13] See Quine (1968).

[14] McGee (2000) shows that the disturbing situation can be resolved for the special case where the substitution instances are proper names by considering counterfactual reasoning. Assume that there is an unnamed individual not living in our world that has a property P and that nobody in our world has P. In a world w where all the inhabitants of our world lives in harmony with our unnamed friend, the sentence $\exists xPx$ is true if quantification is objectual and false if substitutional. As McGee puts it, "Inseparability is an accidental feature of this world, and once we begin looking at other worlds, Quine's problem disappears"(p. 57).

For further discussions, see McGee (2000) and Lavine (2006).

tional seems to make it fairly clear that the generalisations in *Begriffsschrift* are less than absolute. The only further assumption needed is that not everything, objectually understood, is denoted by an expression in $\mathscr{L}_{Bs}$.

2.2 *Grundgesetze*: Russell's paradox and indefinite extensibility

Grundgesetze, it is often claimed, contains an early but unsuccessful account of absolute quantification.[15] The quantifiers therein range over absolutely all objects. Among the objects, Frege counts extensions of concepts which, together with Basic Law V, allows the derivation of Russell's paradox. Scrutinising the semantic underpinnings of *Grundgesetze* and the adoption of Basic Law V, Dummett has argued that Frege's failure consisted in not being aware of the existence of indefinitely extensible concepts. This argument generalises into his highly influential claim that absolute quantification in general, not only in *Grundgesetze*, is untenable because of the indefinitely extensible concepts.

After presenting the argument in Dummett (1991) and the counter-argument in Cartwright (1994) we discuss the implications of these arguments for the possibility of providing a model-theoretic semantics for absolute quantification.

2.2.1 The logical system of *Grundgesetze*

At the level of syntax *Grundgesetze* uses the same formal system as *Begriffsschrift* except for two new primitive symbols: the symbol for the abstraction operator '$\acute{\varepsilon}$' and the symbol for the definite article '\'. In '$\acute{\varepsilon}$' the purpose of 'ε' is to bind occurrences of this sign in the expression that follows '$\acute{\varepsilon}$'. Thus, for instance, in '$\acute{\varepsilon}\Phi(\varepsilon)$', the first occurrence of 'ε' binds the second.

Despite syntactic similarities, the system of *Grundgesetze* is given a radically different semantics than the system of *Begriffsschrift* . One striking difference is that the syntactic distinction in *Begriffsschrift* between function and argument is now paralleled by a corresponding semantic distinction.

[15]One example might be the quote at the beginning of this chapter.

Another, equally important, distinction is that between *Sinn* and *Bedeutung*.

Roughly, the *Sinn* expressed by a sign is the way in which the *Bedeutung* is established, and the *Bedeutung* is what the sign refers to. Having the same *Sinn* entails having the same *Bedeutung*, whereas having the same *Bedeutung* by no means entails that the signs in question share their *Sinn*. Thus, instead of analysing identity statements such as '$\Delta = \Gamma$' as 'the sign Δ has the same content as the sign Γ' Frege now says that the identity statement has as *Bedeutung* the True if, and only if, 'Δ' has the same *Bedeutung* as 'Γ', but this does not entail that 'Δ' has the same *Sinn* as 'Γ'. For Frege, the identity relation is now a relation between objects, not between signs.[16]

The syntactic distinction between function and argument in *Begriffsschrift* appears in *Grundgesetze* as a distinction between function-names and argument-names.[17] Function-names are characterised by being incomplete, or unsaturated, and when completed with appropriate argument-names they become names of objects.[18] Just as function-names are incomplete, so are the functions they denote; functions are incomplete objects, i.e., they are not really objects. An object denoted by a name resulting from the completion of a function-name, Frege calls the value of the function designated by the function-name.

The concepts form a sub-collection of the functions. The value of a concept for any argument, or arguments, is always a truth-value.

Although a function is designated by a function-*name*, we said that it is not counted as an *object*. To be designated is not a sufficient criterion for objecthood:

> *Objects* stand opposed to functions. Accordingly I count as *objects* everything that is not a function, for example, numbers, truth-values, and courses-of-values to be introduced below. The names of objects—the *proper names*—therefore carry no argument-places; they are saturated, like the objects them-

[16] For a detailed defence of this claim, see Heck (2003), and for a discussion of the evolvement of Frege's views on identity, see May (2001).

[17] *Grundgesetze* §1.

[18] *Grundgesetze* §2.

> selves. (*Grundgesetze* §2, italics in the original)

If objects were the only arguments to functions, the function-argument distinction would accomplish a partition of the *Bedeutungs*. However, it is not a necessary condition for arguments to be saturated, and hence, the function-argument distinction does not agree with the function-object distinction. Functions too can be arguments.[19]

Among the primitive functions that take objects as arguments we have $-\!\!-\!\!-\,\xi$ which sends the True to the True and everything else to the False. $-\!\!\!\top\!\!\!-\,\xi$ on the other hand sends the True to the False and everything else to the True. Thus, the True and the False are the only possible values of the functions just considered. Furthermore, the truth-values are conceived of as saturated and hence count as objects.

An example of a function which takes functions as arguments is the (first-order) universal quantifier. Whereas we found strong reasons to understand quantification in *Begriffsschrift* as substitutional it is clearly objectual in *Grundgesetze*. Thus:

> $\vdash\!\!\overset{\mathfrak{a}}{\smile}\!\!-\,\Phi(\mathfrak{a})$ is to denote the True if for every argument the value of the function $\Phi(\xi)$ is the True, and otherwise is to denote the False, [...] (§8)

In '$\vdash\!\!\overset{\mathfrak{a}}{\smile}\!\!-\,\Phi(\mathfrak{a})$', the function-name '$\Phi(\xi)$' may be considered a mark of an argument-place.[20] Thus, '$\vdash\!\!\overset{\mathfrak{a}}{\smile}\!\!-\,\Phi(\mathfrak{a})$' is a mark of a function which takes functions as arguments and it denotes a function which sends every function which is true for every argument to the True.

Now, according to Frege, a function that takes objects as arguments is of another kind than a function that takes functions as arguments. From §21 it seems clear that at least one aspect of this matter rests upon syntactic considerations. In place of '$\Phi(\xi)$',

> [...] only names of functions of one argument—not proper names, nor names of functions of two arguments—may be

[19] §19–§24.

[20] Just as in *Begriffsschrift*, capital Greek letters occur in the general outline of the system as meta-variables which on each occasion must be thought of as having a definite value.

> substituted, for the combinations of signs being substituted must always have open argument-places to receive the letter "𝔞", and if on the other hand we wanted to substitute a name of a function of *two* arguments, then the ζ-argument-places would remain unfilled. (§21)

Hence there are syntactical reasons to separate different types of functions, at least when they appear as arguments.[21] But as Frege's syntax is meant to fully reflect the structure of the underlying semantics the syntactic distinctions correspond to distinctions at the level of semantics. Functions that take objects (and only objects) as arguments are called *first-level functions*, functions taking first-level functions as arguments are *second-level functions*, and so forth. The arguments are then divided into types:

- *arguments of type 1*: objects
- *arguments of type 2*: first-level functions of one argument
- *arguments of type 3*: first-level functions of two arguments[22]

Corresponding to this semantic hierarchy, Frege also divides the argument-places into types in an analogous syntactic hierarchy. Thus, for example, the universal quantifier is a second-level function of arguments of type 2. Frege also defines quantification over first-level functions of one argument. Such a quantifier is a third-level function of second-level functions.

This embryo of a theory of types is interesting for at least two reasons. First, it is not motivated by the paradoxes as many of the subsequent type theories, e.g. the theory of types in *Principa Mathematica* (Russell and Whitehead, 1910). This may indicate that type theories are less ad hoc than one might think.[23] Secondly, it seems to imply that Frege did not understand his quantifiers as ranging over absolutely everything; since no quantifier in the formal system of *Grundgesetze* ranges across different types, it follows that no quantifier ranges over both objects and functions. In other words, generalisations are always confined to one, and only one, of

[21] §23.
[22] Frege does not consider functions of more than two arguments.
[23] See Maddy (1997, ch. 1) for a discussion.

the several mutually exclusive types. We will have reason to return to this observation in Section 5.1, when discussing a semantics proposed in Williamson (2003).

Falling short of quantifying over absolutely everything, Frege clearly took some of his quantifiers as ranging over all *objects*. Interestingly, Dummett's analysis of the inconsistency in *Grundgesetze* already makes quantification over all objects doubtful. Then, if Dummett is right and we cannot quantify over all Fregean objects, it seems unlikely that we should be able to quantify over the possibly greater totality of everything there is.

Besides the truth-values, Frege counts *courses-of-values* among the objects. If $\phi(\xi)$ is a first-level function of one argument, then the *course-of-values* of $\phi(\xi)$ is denoted by $\acute{\varepsilon}\phi(\varepsilon)$. The identity criteria for *courses-of-values* is the Basic Law V:

(8) $\vdash\!\!-\!\!-\!\!- (\acute{\varepsilon}f(\varepsilon) = \acute{\alpha}g(\alpha)) = (-\!\overset{\mathfrak{a}}{\smile}\!- f(\mathfrak{a}) = g(\mathfrak{a}))$

This axiom settles the denotations for identity statements for *courses-of-values* as long as each *course-of-values* is given on the form $\acute{\varepsilon}\phi(\varepsilon)$. However, in §10, Frege raises the question if we can recognise a *course-of-values* as such if it is *not* given on the form $\acute{\varepsilon}\phi(\varepsilon)$. This is crucial in order to sort out the truth-value of functions like $\acute{\varepsilon}f(\varepsilon) = \xi$ for different arguments.

Dummett notes that Frege's solution in §10 is based on a context principle saying that a singular term has a reference only if "the result of inserting it into the argument-place of any functional expression of the language has a reference."(Dummett, 1991, p. 212) Frege shows that it is enough to consider identity statements since the other functions reduce to such statements in the relevant cases. Furthermore, since the only objects introduced up to that point are the truth-values the question boils down to whether these objects may be identified as particular *courses-of-values*. Thus, in §10, Frege identifies the True with $\acute{\varepsilon}(-\!\!-\!\!-\!\!-\varepsilon)$, and the False with $\acute{\varepsilon}((\vdash\!\!\top\!\overset{\mathfrak{a}}{\smile}\!- \mathfrak{a} = \mathfrak{a}) = \varepsilon)$. This solves the problem of determining the value of $\acute{\varepsilon}f(\varepsilon) = \xi$ for all arguments.

To get the result that a term for a *course-of-values* designates some object Dummett recognises that Frege also tacitly assumes a principle of compositionality for having a reference:

> [...], if the result of inserting a term into the argument-place of every *primitive* functional expression has a reference, then the result of inserting it into the argument-place of *any* functional expression will have a reference. We may call this the 'compositional assumption'. (Dummett, 1991, p. 212, italics in the original)

Frege uses this principle also in §§29–31 when arguing that each proper name, and each first-level function of the language in *Grundgesetze* has a denotation.[24]

We now proceed to the derivation of Russell's paradox in *Grundgesetze*.

2.2.2 Russell's paradox

Frege gives two formal derivations and discusses Russell's paradox in Appendix II of the second volume of *Grundgesetze*. We follow Frege's first derivation using a slightly modernised terminology.

Call an extension a *class* whenever it is an extension of a concept. The concept *class of all classes not belonging to themselves* is designated by means of

(9) $\neg\forall G(\acute{\varepsilon}G(\varepsilon) = \xi \rightarrow G(\xi))$

Its extension is designated by:

(10) $\acute{\alpha}(\neg\forall G(\acute{\varepsilon}G(\varepsilon) = \alpha \rightarrow G(\alpha))$

By the preceding discussion (10) has a denotation. Let W abbreviate (10). Using Basic Law V, from left to right, we obtain

(11) $\acute{\varepsilon}f(\varepsilon) = W \rightarrow (f(W) \leftrightarrow \neg\forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W)))$

That is, this follows by instantiation from the left-to-right direction of

$$\acute{\varepsilon}f(\varepsilon) = \acute{\alpha}g(\alpha) \leftrightarrow \forall x(f(x) \leftrightarrow g(x))$$

where $\neg\forall G(\acute{\varepsilon}G(\varepsilon) = \alpha \rightarrow G(\alpha))$ is substituted for $g(\alpha)$.

[24] See also Dummett (1991, pp. 209–216).

Note that, if W is not a class, i.e. if it were an object other than a *course-of-values*, (11) would involve the Julius Caesar problem.[25] What bars this problem here is the austere ontology in *Grundgesetze*, which, by the context principle and the principle of compositionality, secures that (11) denotes a truth-value.[26]

The rest of the derivation is straightforward. From (11) and propositional logic we have

$$(12)\ \neg\forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W)) \rightarrow (\acute{\varepsilon}f(\varepsilon) = W \rightarrow f(W))$$

which in turn, by second-order generalisation, gives

$$(13)\ \neg\forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W)) \rightarrow \forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W))$$

Next, an instance of second-order instantiation is

$$(14)\ \forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W)) \rightarrow (\acute{\varepsilon}f(\varepsilon) = W \rightarrow f(W))$$

which by substituting (9) for $f(\xi)$, together with the definition of W, gives

$$(15)\ \forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W)) \rightarrow \neg\forall G(\acute{\varepsilon}G(\varepsilon) = W \rightarrow G(W))$$

Now (13) and (15) yields the contradiction.

Frege concludes that the only possible error lies in Basic Law V and that

> [...] we must take into account that possibility that there are concepts having no extension—at any rate, none in the ordinary sense of the word. Because of this, the justification of our second-level function $\acute{\varepsilon}\psi(\varepsilon)$ is shaken; yet such a function is indispensable for laying the foundation of arithmetic. (Frege, 1893,1903, pp. 131–132)

[25] The Julius Caesar problem, as applied here, consists in the problem of deciding the denotation, i.e. the truth-value, of an identity statement $\acute{\varepsilon}f(\varepsilon) = W$ where W is not given as a *course-of-values*. In that case the right hand side of Basic Law V doesn't determine its denotation.

[26] See §31.

2.2.3 Dummett's argument and the All-in-One Principle

The non-ordinary sense of the word 'extension' that Frege speaks of in the preceding quote is explained by Dummett in terms of indefinite extensibility:

> It is clear that Frege's error did not lie in considering the notion of an extension of a concept to be a logical one, for that it plainly is. Nor did it lie in his supposing every definite concept to have an extension, since it must be allowed that every concept defined over a definite totality determines a definite subtotality. We may say that his mistake lay in supposing there to be a totality containing the extension of every concept defined over it; more generally, it lay in his not having the glimmering of a suspicion of the existence of indefinite extensible concepts. (Dummett, 1991, p. 317)

A concept is indefinitely extensible if its having a definite extension gives rise to new instances of the concept in question. Thus, for instance, the concept of *non-self-membered class* qualifies as an indefinitely extensible concept. For assume it had an extension R, then, on pain of contradiction, R cannot be a member of R, which makes it a new instance of the concept.[27]

Clearly, elucidations as the above do not fully characterise the concept of indefinite extensibility. For one thing, explaining *indefinite* extensibility in terms of having a *definite* extension seems dangerously circular. In fact, characterising indefinite extensibility has turned out quite a challenge that, however, we will not delve into here.[28] For our purposes, it will be enough to assume that the explanation above is sufficiently clear.

Dummett argues that in order to "obtain a determinate interpretation of a formal language we must first specify, without circularity, what the elements of the domain are to be, before we go on to specify the intended interpretations of the primitive predicates [...]" (Dummett, 1991, p. 221). Frege did not specify such a domain. Instead, as Dummett points out, he

[27] Dummett (1991, p. 317). Indeed, Russell expressed similar ideas in Russell (1907).
[28] See Shapiro and Wright (2006) for a rewarding discussion.

tried to fix the references of names of objects by means of an ill-founded application of the context principle. Thus,

> [a]lthough there is in fact no danger of inconsistency in the fragment of Frege's system with only first-order quantification, he has provided no valid proof of its consistency, because he has not succeeded in specifying the references to all its terms. For that reason, he failed to justify the introduction of value-ranges. (Dummett, 1991, p. 222)

Since the first-order fragment is consistent it has a model. Let us for the sake of argument assume that the domain of this model, i.e., the domain over which the first-order quantifiers range, encompasses absolutely all objects. Second-order quantification gives us the means to construct a determinate concept such as (9). The extension of (9) as defined over the domain of all objects may instantiate Basic Law V and, thus, gives rise to a contradiction. What the argument shows, according to Dummett, is thus that the extension of (9) wasn't among the objects in the first place. That is, it wasn't within the range of the first-order quantifiers. For if it were, since (9) is definite according to Frege, we have the contradiction. It follows that the domain that supposedly contained all objects could be extended and, hence, that the first-order quantifiers didn't range over all objects after all.

One line of critique, articulated by Cartwright (1994), of using this type of argument against unrestricted quantification, claims that it rests upon a false principle, viz. the All-in-One Principle:

> [...] to quantify over certain objects is to presuppose that those objects constitute a "collection," or "completed collection"—some one thing of which those objects are the members. (Cartwright, 1994, p. 7)

Cartwright's argues that this principle is false in a way that is reminiscent of Boolos (1984, 1985):

> There would appear to be every reason to think it false. Consider what it implies: that we cannot speak of the cookies in the jar unless they constitute a set; that we cannot speak of

> the natural numbers unless there is a set of which they are the members; that we cannot speak of all pure sets unless there is a class having them as members. I do not mean to imply that there is no set the members of which are the cookies in the jar, nor that the natural numbers do not constitute a set, nor even that there is no class comprising the pure sets. The point is rather that the needs of quantification are already served by there being simply the cookies in the jar, the natural numbers, the pure sets; no additional objects are required. (Cartwright, 1994, p. 8)

What the set-theoretic paradoxes may have revealed is the existence of indefinitely extensible concepts, and thus, that there cannot be a totality of everything that might be a possible value of our first-order variables, i.e., no universal domain. But without the All-in-One Principle, Cartwright argues, the general argument is not conclusive. Thus,

> Dummett's argument seems to be simply that since there is no universal "domain," and since the All-in-One Principle is true, unrestricted quantification is illegitimate. (Cartwright, 1994, p. 17)

In the next section we discuss what consequences the falsity (or truth) of the All-in-One Principle might have for the possibility of providing a model-theoretic semantics for absolute quantification.

2.2.4 The All-in-One Principle and model-theoretic semantics

Here is a natural but unsound argument that the adoption of a model-theoretic semantics entails the truth of the All-in-One Principle: A model $\mathcal{M}$ of a first-order language $\mathscr{L}$ is an ordered pair $\langle M, I \rangle$ of a set M to which the quantifiers of $\mathscr{L}$ are restricted and a function I interpreting the non-logical vocabulary. Thus, since the quantifiers in $\mathscr{L}$ are always restricted to some set of the model interpreting $\mathscr{L}$, quantification always presupposes, in a model-theoretic semantics, that the objects quantified over constitute a set, or a completed collection. Hence the All-in-One Principle holds.

The above argument is not conclusive since it fails to distinguish between the use of quantifiers and the mentioning of them, e.g. in semantic theorising. It seems clear enough, as Cartwright points out, that when we *use* our quantifiers, all we have to acknowledge as existing are the entities quantified over, not the collection of them as an additional entity. This is true even if we adopt a model-theoretic semantics in which we do quantify over domains of quantification. For the All-in-One Principle concerns the use, rather than the mention of quantifiers in theories, and as such it seems to be a false principle.

The situation is slightly different when a quantifier is used to express absolute generality. For if we want to interpret such a quantifier faithfully, whatever is within the range of the quantifiers of the metalanguage, e.g., domains of quantification, must also be within the range of the interpreted quantifier. In a way then, the quantifiers in the object language inherit the ontological commitment attached to the quantifiers in the metalanguage. Thus we might think that, given a model-theoretic semantics, any use of quantifiers to express absolute generality makes us committed to domains of quantification after all. At least this seems to be the case if such domains are quantified over in the metatheory. Since quantification over domains is necessary for an adequate definition of logical consequence, one might argue that the adequacy of model-theoretic semantics seems to entail the All-in-One Principle after all.

But this would be a mistake. For, typically, in semantic theory, we use certain set-theoretic constructions to *represent* other entities, but that does not mean that the entities represented *are* those set-theoretic constructions. For instance, just because we standardly represent the natural numbers $0, 1, 2, \ldots$ by the sets $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots$, we do not identify the natural numbers with those sets. The same holds for domains of quantification. Hence, rather than domains of quantification, the inherited ontological commitment in the object language is of the representatives of such domains, i.e., in case of model-theoretic semantics, sets.

Thus, when quantifying over absolutely everything in an object language, we are not committed to domains of quantification, just because we say that we quantify over such domains in the metalanguage, for what we really quantify over are the representatives of such domains. Thus, model-theoretic semantics imposes in such an object language commitment to sets rather than sets-as-domains of quantification. Hence, model-theoretic semantic does not support the All-in-One Principle just by using sets to represent domains of quantification in its models.

One may, at this point, wonder why we should bother with the truth of the All-in-One Principle at all. We saw that, according to Cartwright, the All-in-One Principle, together with the impossibility of a universal domain of quantification, constitutes the core of Dummett's argument against the coherence of absolute quantification. Given the impossibility of a universal domain, the strength of Dummett's argument hinges on the truth or falsity of the All-in-One Principle. But we will argue that the impossibility of a universal domain of quantification is far from given. On the contrary, our ultimate aim in Chapter 6 is the construction of a model-theoretic semantics in a set theory that has a universal set playing the role of such a domain. Thus Dummett's argument is blocked regardless of whether the All-in-One Principle turns out to be true or false. Hence, once we have seen that model-theoretic semantics does not entail the truth of the All-in-One Principle by the argument above, that principle may seem irrelevant for our purposes.

But then again, the All-in-One Principle might still turn out to be relevant for a model-theoretic semantics for absolute quantification. For, as Cartwright notices, "the All-in-One Principle may be thought to derive support from current model-theoretic accounts of first-order logical truth and consequence." (Cartwright, 1994, p. 11)

To explain the idea Cartwright uses the notion of a *set-theoretic analogue* of a logical schema[29]. The set-theoretic analogue of a formula φ of some language $\mathscr{L}$ is reached in two steps: first, uniformly assign variables y to each predicate letter P of $\mathscr{L}$, and then replace each occurrence of $P(x_0, \ldots, x_n)$ in φ by $\langle x_0, \ldots, x_n \rangle \in y$.

It is tempting, Cartwright says, to identify the logical truths with those sentences whose set-theoretic analogues have true universal closures. That would amount to saying that "a sentence of a first-order language is a logical

[29]The notion comes from Quine (1970), see in particular p. 51.

truth just in case it is true on any assignment of extensions to its predicates [...]." (Cartwright, 1994, p. 9) But, Cartwright continues, this would lead us astray, e.g. by implying that $\neg\forall x Px$ is a logical truth. For its set-theoretic analogue is the formula $\forall y \neg\forall x(x \in y)$ which is true according to standard set theory. Thus, since $\neg\forall x Px$ is certainly not a logical truth, the identification between logical truths and sentences with true universal closures of their set-theoretic analogues is mistaken.

It is perhaps telling that this counterexample is invalid in set theories with a universal set. Indeed, it is not at all straightforward to tell whether there are such counterexamples for the alleged identification in, for instance, Quine's set theory NF, or $\mathrm{NFU_p}$, the theory we will eventually use as our metatheory.

But Cartwright makes his point in relation to standard set theories where the counterexample works. He continues to note that what is missing in the definition of logical truth by means of set-theoretic analogues is the relativisation to universes. The relativisation to a universe U of a set-theoretic analogue of some formula φ is simply the set-theoretic analogue with its quantifiers restricted to appropriate constructs on U. Thus, the quantifiers of φ will be restricted to U, and, taking the polyadicity of the predicate letters into account, the remaining quantifiers of the set-theoretic analogue are restricted to sets of pairs, triples, etc. of elements of U. Logical truth is then defined as truth of the set-theoretic analogue under all relativisations.

Returning to the counterexample, we see that the relativised set-theoretic analogue of $\neg\forall x Px$ now becomes

$$\forall U \forall y(y \subseteq U \to \exists x(x \in U \land x \notin y)),$$

which is easily seen to be false: pick a U and put $y = U$. Thus the set-theoretic analogue of $\neg\forall x Px$ is not true under all relativisations, and the sentence is not a logical truth.

Of course, defining logical truth as truth of the set-theoretic analogues under all relativisations is just to define it as truth in all structures of the right signature. The point, however, of taking the detour via the unrelativised set-theoretic analogues, to reach that familiar definition, is to show the necessity of universes. That necessity may, in turn, be taken to support the All-in-One Principle in the following way:

> [T]he amended definition may encourage the idea that among the relativizations of the universal closures of the set-theoretic analogues of the first-order sentence φ there must be one that expresses the intended interpretation of φ, or at least a proposition equivalent in some strong sense to the intended interpretation of φ. It might be thought that otherwise truth of the relativizations would give no guarantee even of the plain *truth* of φ.[30] (Cartwright, 1994, p. 10, italics in the original)

Thus, given that the intended interpretation ought to be, or correspond closely to, one of the relativised set-theoretic analogues, then it seems as if the intended interpretation necessarily involves a universe of quantification, in this case represented by a set. Thus, we seem to have an argument for the All-in-One Principle from the model-theoretic account of logical truth.

Notice that this might turn out a problem regardless of which set theory we use to construct our model theory. That model-theoretic semantics supports a false principle is potentially a problem in itself. Indeed, if we are prepared to take 'support' to mean something close to *entail*, we would have a reductio for the adequacy of logical truth and consequence as defined in model-theoretic semantics.

But, Cartwright continues, due to the completeness theorem for first-order logic, the above argument is not binding. For, clearly, provability of a sentence φ entails the plain truth of φ, and if the set-theoretic analogue of this sentence is true on all relativisations, then it follows from completeness that it is provable. Hence, the logical truth of a sentence entails its plain truth even if the intended interpretation of it is not among its relativised set-theoretical analogues.[31]

After this lengthy discussion, my general conclusion is that if we choose to adopt a model-theoretic semantics for absolute quantification we are not thereby automatically committed to the All-in-One Principle.

---

[30] Cartwright uses p instead of φ.

[31] Of course this is Kreisel's (1967) famous squeezing argument, which, in other contexts, has received a lot of attention recently; see e.g. Smith (2011). Cartwright reports that his version is derived from Quine (1970, pp. 54–55).

3 Tarski's Definition of Truth and Logical Consequence

As all students of logic and philosophy know, Tarski gave the first mathematically acceptable definition of truth in his monograph *The Concept of Truth in Formalized Languages* (CTFL).[1] Despite this there seems to be, as Feferman (2008, p. 76) puts it, a "*prima facie* discrepancy between what logicians nowadays usually say Tarski did, and what one finds in the *Wahrheitsbegriff* [CTFL]." What one does *not* find was pointed out by Wilfrid Hodges:

> A few years ago I had a disconcerting experience. I read Tarski's famous monograph 'The concept of truth in formalized languages' [Tarski (1935)] to see what he says himself about the notion of truth in a structure. The notion was simply not there. This looked curious, so I looked in other papers of Tarski. [...]
>
> I believe that the first time Tarski explicitly presented his mathematical definition of truth in a structure was his joint paper with Robert Vaught [Tarski and Vaught (1957)]. (Hodges, 1985/6, p. 137–138)

Hodges asks why Tarski didn't define the concept of *truth in a structure* in CTFL. After all, by then he seems to have had all the tools needed for such a definition.[2]

[1] CTFL was first published in Polish in 1933, though most of the results stem from 1929. It was translated to German in 1935 and to English in 1956. All references made here are to the English translation Tarski (1935).

[2] See Hodges (1985/6, p. 138). Feferman (2008) claims that Tarski "worked comfortably with the informal notion of model for first-order and second-order languages at least since 1924."

Part of an answer to Hodges's question lies in the pre-theoretical concept of truth that Tarski sets out to define in CTFL. Here is a telling passage:

> Amongst the manifold of efforts which the construction of a correct definition of truth for the sentences of colloquial language has called forth, perhaps the most natural is the search for a *semantical definition*. By this I mean a definition which we can express in the following words:
>
> (1) *a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.* (CTFL p. 155. Italics in the original.)

This kind of truth definition aims at defining truth for meaningful languages, i.e., languages that contain sentences that actually say that states of affairs do or don't uphold. Tarski regards the problem of defining the semantic concept of truth for purely formal languages as irrelevant, or even meaningless, since sentences of formal languages do not express something determinate about reality and therefore cannot, in the semantic sense, be true or false.[3] Hence Tarski sets out to define truth for languages whose signs have, as he puts it, "intelligible meanings" (CTFL, p. 167) which excludes purely formal languages. This explains why Tarski did not define the concept of truth in a structure in CTFL. For, if only meaningful languages are considered, no structures are needed for sentences to have contents, and hence the concept of *truth in a structure* is simply uncalled-for.

Structures, or models, standardly restrict quantification to a set—the domain of quantification—and are usually considered to be inept to interpret absolute quantification. That Tarski doesn't make use of structures to define truth is thus hardly a loss from the perspective of absolute quantification. On the contrary, CTFL provides a context in which it is rather natural to formulate questions regarding absolute quantification. For instance, given that a definition of the semantic concept of truth is relevant only for languages that are meaningful in the sense of somehow being about reality, one may ask if there are among these languages those for which we may construct a truth definition according to which there are true sentences

[3] See CTFL, p. 166–7. See also Hodges (1985/6, p. 148).

about the whole of reality, i.e., true sentences which are about absolutely everything.

In Section 3.1 this question is discussed with respect to the general framework in CTFL, and we give a negative answer. In Section 3.2 two natural proposals of how to tweak the definition to allow for absolute quantification are suggested and briefly discussed. One of these suggestions lies close to type-theoretic semantics that we discuss further in Chapter 5. The other suggestion uses set-theoretic constructions in the definition. In Section 3.3, we show that the resulting definition of truth has unwanted implications for the concept of logical consequence. At least it has such consequences if we require that the membership relation is well-founded.

3.1 Absolute quantification and truth in CTFL

Tarski considers at some length in CTFL the feasibility of constructing an adequate definition of truth for natural language but concludes that such a definition is problematic in at least two ways. One problem is the apparent lack of a firm structure in natural language:

> [Natural] language is not something finished, closed, or bounded by clear limits. It is not laid down what words can be added to this language and thus in a certain sense already belong to it potentially. We are not able to specify structurally those expressions of the language which we call sentences, still less can we distinguish among them the true ones. (CTFL, p. 164).

This lack of structure is problematic for Tarski since he sets out to define *true sentence* ultimately in purely structural, non-semantic terms.[4] And indeed, if natural language is like Tarski says it is, a structural specification of *sentence*, and a fortiori, a definition of *true sentence*, seems remote.

The second problem that Tarski identifies with natural language is, what he calls, the property of universality:

> A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality. It would not be

[4] See CTFL, p. 153.

> in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language'. [...] it is presumably just this universality of everyday language which is the primary source of all semantical antinomies, like the antinomies of the liar or of heterological words. (CTFL, p. 164)

Tarski famously avoids the semantic paradoxes by developing the semantics for a language, the object language, in another language, the metalanguage. The metalanguage has to be systematically separated from the object language in order for this distinction to function as a guard against the semantic paradoxes. The main problem with universality then seems to be that it blurs this distinction and, hence, spoils Tarski's guard against the paradoxes.

Typically, a metalanguage of an object language possesses resources to speak about the syntax of the object language and contains translations of the expressions of the object language. In particular, assuming that 'true sentence' has been defined, the metalanguage contains Tarski's T-sentences, i.e., sentences of the form

(T) x is a true sentence if and only if p,

where x is a name of a sentence in the object language and p is the translation of that sentence in the metalanguage.[5] In fact, a definition of truth is said to be *adequate* only if all instances of (T) follow from it.[6] Now, if an object language possesses the property of universality, all that can be said in the metalanguage can also be said in the object language and, in particular, given an adequate definition of truth, all instances of the T-schema, which normally belong only to the metalanguage, are available already in the object language.[7] Thus, the distinction between the languages collapses

[5] Tarski does not use the name 'T-sentences' in CTFL, but see Tarski (1944).
[6] This is one part of convention T, CTFL, p. 187. See also p. 46.
[7] See CTFL, p. 167.

and, given some natural assumptions, the liar paradox becomes derivable.[8] Hence, the assumption of an adequate definition of truth for a language with universality leads to inconsistency.

Tarski concludes that

> [...]*the very possibility of a consistent use of the expression 'true sentence' which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression.* (CTFL, p. 165. Italics in the original.)

Consequently, no further efforts of defining truth for natural language are made in CTFL. Instead Tarski turns to *formalised languages*.

A formalised language is explained as the result of a formalisation, i.e., the replacing of a language with an imprecise syntax, or part of such a language, by a language with a precise syntax, but which, in other aspects, diverge as little as possible from the language formalised.[9] In particular, a formalised language contains a vocabulary from which composite expressions may be formed by means of structural formation rules. The sentences form a distinguished sub-category of the expressions.

A precise and complete characterisation of the notion of *formalised language* in CTFL is a matter of some delicacy and we shall not attempt to give one here.[10] For our purposes we may think of a formalised language as a first- or higher-order language (in its modern sense) where the signs used, save for the variables, have fixed meanings. It is important to note, however, that a formalised language, as opposed to a formal one, is meaningful

[8] Let '*c*' name '*c* is not a true sentence' so that *c* = '*c* is not a true sentence' and instantiate (T) to get:

> '*c* is not a true sentence' is a true sentence if and only if *c* is not a true sentence.

Using the identity we get the contradiction:

> *c* is a true sentence if and only if *c* is not a true sentence.

[9] See Tarski (1944, p. 347).

[10] See CTFL, p. 166. For a discussion of the slightly wider question of what languages have truth definitions of the type Tarski constructs in CTFL, see Hodges (2004).

on its own. No interpretation, or structure, is needed to give meanings to the non-logical vocabulary and the sentences are regarded simply as being true or false, rather than true or false in this or that interpretation.

Since no interpretations are needed to give meaning to the languages considered in CTFL, the concept of truth therein is not relative to interpretations in the way we are used to from model-theoretic semantics. But it is nevertheless a relative concept since each definition of truth is given for a particular object language. Thus, one may say, rather than being relative to interpretations, the concept of truth in CTFL is relative to languages. And just as there is a multiplicity of languages, there is a parallel multiplicity of truth definitions. This makes Tarski hesitant about giving a detailed general abstract description of how to define truth for arbitrary languages.[11] Instead he gives a detailed example of a definition of truth for the particular language, $\mathscr{L}_c$, of the calculus of classes. He trusts the reader to make the necessary changes and amendments if some other language of a similar kind would be considered.[12]

We briefly sketch Tarski's definition of truth for $\mathscr{L}_c$ in a metalanguage, $\mathscr{L}_c^m$. $\mathscr{L}_c$ is one sorted and besides the variables $x, y, z, \ldots$, it contains the sign for negation 'N', disjunction 'A', universal quantification 'Π', and inclusion 'I'. Disregarding the variables, these symbols are translated into $\mathscr{L}_c^m$ as 'not', 'or', 'for all' and 'is included in' or '$\subseteq$', respectively. Complex expressions of $\mathscr{L}_c$, and in particular sentences, is formed in accordance with the formation rules of Polish notation and have straightforward translations into $\mathscr{L}_c^m$. Thus, for instance, 'I*xx*', 'NAI*xy*I*zx*' and 'Π*x*I*xx*', are well formed expressions that translate into '$x \subseteq x$', 'not ($x \subseteq y$ or $z \subseteq x$)', and 'for all x, $x \subseteq x$'.[13]

Besides translations of the expressions in $\mathscr{L}_c$, $\mathscr{L}_c^m$ contains what Tarski calls structural descriptive names of those expressions, that is, names that mirror the syntactic structure of the expressions named. To simplify the

[11] See Tarski (1935, pp. 167–168).

[12] See Tarski (1935, pp. 209–210).

[13] We avoid indexing the variables in the metalanguage if possible. Also we let the same signs $x, y, z, \ldots$ be variables in both languages. We sometimes even let the variables figure as names of themselves. In what role a variable appears should always be clear from the context.

exposition we use φ and ψ as variables ranging over expressions in $\mathscr{L}_c$. We also anachronistically allow mixed expressions such as '$\Pi x \varphi$' as part of $\mathscr{L}_c^m$.

The definition of truth uses infinite sequences which Tarski identifies with one-many relations having the set of variables as their converse domain. That is, given a class in the domain of a sequence f, it may bear the relation f to any number of variables. By f_x we mean the class that is related by f to x.

With the sequences in place, the definition of truth goes via the concept of *satisfaction* which is defined by recursion on the complexity of the expressions of $\mathscr{L}_c$. The detour via satisfaction is necessary since truth is applicable only to sentences and the components of sentences include sentential functions which are neither true nor false.

We formulate the definition of satisfaction for $\mathscr{L}_c$:[14]

Definition 3.1.1. Let φ, ψ and χ be sentential functions in $\mathscr{L}_c$. Then, the sequence f *satisfies* the sentential function φ if and only if f is an infinite sequence of classes and φ is a sentential function such that either

1. φ is of the form Ixy and $f_x \subseteq f_y$, or
2. φ is of the form Nψ and f does not satisfy ψ, or
3. φ is of the form A$\psi\chi$ and f satisfies ψ or f satisfies χ, or
4. φ is of the form $\Pi x \psi$ and every infinite sequence of classes that is like f, except possibly for f_x, satisfies ψ.

A sentence of $\mathscr{L}_c$ is now said to be *true*, or a *true sentence*, if it is satisfied by all sequences in accordance with the above definition.

Thus defined truth concerns $\mathscr{L}_c$-sentences, and clearly, if some other language were considered, we would have to define truth, or rather satisfaction, differently. Were there, for instance, additional primitive vocabulary in $\mathscr{L}_c$, the number of clauses similar to the one for I, would multiply. For instance, if '$\in$' were added as a new primitive symbol to $\mathscr{L}_c$, we would need an additional clause governing the satisfaction of sentential functions such as $\in xy$ in the definition. Thus, as we said, the semantic concept of truth

[14] This definition is given in CTFL, p. 193. We do not use Tarski's original terminology.

that Tarski defines is relative to the object language to which it is meant to apply.

We also said that a definition of truth is adequate if it implies all instances of the T-schema. This is the essence of Tarski's convention T which is officially stated for the particular example of defining truth for $\mathscr{L}_c$:

> Convention T. *A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an* adequate definition of truth *if it has the following consequences*:
>
> (α) *all sentences which are obtained from the expression '$x \in Tr$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage*;
>
> (β) *the sentence 'for any x, if $x \in Tr$ then $x \in S$' (in other words '$Tr \subseteq S$').* (Tarski, 1935, p. 187, italics in the original.)

S is understood as the collection of sentences of the object language considered.

Even though convention T is stated in the particular context of defining truth for $\mathscr{L}_c$, it is clearly meant to be a general convention.[15] Any semantic definition of truth, for any object language, must allow for the deduction of each instance of (α) and (β) in order to be adequate. The generality results from a certain amount of indeterminacy. If one reads the convention as its stands, it is by no means clear, in the general case, what metatheory is to be employed to deduce the T-sentences as consequences of the definition of *Tr*.[16] In the particular case of $\mathscr{L}_c$ this poses no troubles since the metalanguage in which the definition of *Tr* is carried out is explicitly stated in §2. But generally, given an object language, in order to make convention T determinate, some metalanguage has to be provided.

Tarski discusses in §§4–5 if there is always a suitable metalanguage for each possible object language. That is, if there is, for any given object lan-

[15] For example, see CTFL, p. 246, and Tarski (1944, p. 344).

[16] See David (2008), especially the section named *The double life of convention T*, for a discussion.

guage, a metalanguage in which an adequate definition of truth for that language can be carried out. A related question of particular interest for us, though perhaps not for Tarski, is to consider the possibility of a metalanguage allowing for a Tarski style truth definition for an object language with absolute quantification.

A quick glance at the definition of satisfaction for $\mathscr{L}_c$ indicates that the situation, at least initially, looks quite promising. For, though Π is explicitly restricted to classes in the fourth clause of that definition, this seems like an incidental feature of the definition. Prima facie it seem quite possible to construct a truth definition for absolute quantification by dropping the restriction of the quantifier, replacing the equivalent of *4* in 3.1.1 with

4′. *φ is of the form $\Pi x\psi$ and every infinite sequence that is like f, except possibly for f_x, satisfies ψ.*

To show that this is a successful method we need to carry out the details of at least one truth definition for at least one object language with absolute quantification. Thus, to begin with, we need an object language in which we quantify over absolutely everything. Also, having such an object language in place, we need to state explicitly what metalanguage we use to construct the definition of truth. But here problems emerge. According to Tarski, all languages obey a variant of Husserl's theory of *semantic categories*. This, in fact, implies that there cannot be a suitable metalanguage in which a definition of truth is possible for a language with absolute quantification.

According to Tarski's version of the theory of semantic categories, two expressions are said to belong to the same semantic category if, in any sentential function containing one of the expressions, the other expression is substituted for this expression, the sentential function remains a sentential function.[17] Consider, for example, 'I*xy*'. Since substituting '*x*' for '*y*' results in the well-formed 'I*xx*', and similarly for any other (free) occurrence of '*x*' in a sentential function, it follows that '*x*' and '*y*' belong to the same semantic category. But, since 'I*x*I' is not well-formed, 'I' and '*y*' belong to different semantic categories.

Husserl does not characterise the semantic categories in terms of sentential functions. Instead he uses the notion of meaningfulness, saying

[17] CTFL, p. 216.

that two expressions belong to the same semantic category if a meaningful expression containing the one expression as (proper) part, remains meaningful if the other expression is substituted for it.[18] Presumably Tarski's characterisation in terms of sentential functions is motivated by his pledge not to use semantic terms in the course of defining truth.[19]

The semantic categories together with their elements are divided into orders and types in a way similar to the simple theory of types in *Principia Mathematica*. But though Tarski regards the concept of semantic category as close to the simplified concept of logical type of *Principia Mathematica*, he also points out that the two theories differ in important aspects with regard to their origin and content:

> Whilst the theory of types was thought of chiefly as a kind of prophylactic to guard the deductive sciences against possible antinomies, the theory of semantical categories penetrates so deeply into our fundamental intuitions regarding the meaningfulness of expressions, that it is scarcely possible to imagine a scientific language in which the sentences have a clear and intuitive meaning but the structure of which cannot be brought in harmony with the above theory. (CTFL, p. 215)

Tarski uses 'scientific language' as a synonym to 'formalised language' and it is clear from this passage that, not only is the theory of semantic categories differently motivated than the theory of types, it also "penetrates so deeply into our fundamental intuition" that all formalised languages must obey this theory. In particular, all object languages and metalanguages in CTFL adhere to the theory of semantic categories.

While Tarski discusses at great length problems that the semantic categories bring about for the definition of truth for higher-order languages in §§4–5 of CTFL, a related problem for absolute quantification appears already in the first-order case. To see this, consider a dyadic relation W and suppose we want to express that

'$\Pi x\, Wxx$' is a true sentence.

[18] See Simons (2001).

[19] As a result, his semantic categories are really syntactic.

For instance, W may be a universal linear ordering, i.e., W may be antisymmetric, transitive and total, and we might want to conclude that W is also reflexive.

As usual a sentence is true if there is a sequence satisfying it. Thus,

> '$\Pi x\, Wxx$' is a true sentence if, and only if, it is satisfied by an infinite sequence f.

Assume for simplicity that the metalanguage in this example contains a copy of the object language. In accordance with $4'$ we then have that

> '$\Pi x\, Wxx$' is a true sentence if, and only if, every infinite sequence that is like f, except possibly for f_x, satisfies Wxx.

Corresponding to the first clause of 3.1.1 we have that

> f satisfies 'Wxx' if, and only if Wf_xf_x.

Now, since we quantify over absolutely everything in the object language there can be nothing over which we quantify in the metalanguage that is not already quantified over in the object language. That is, whatever we quantify over in the metalanguage, we also quantify over in the object language. In particular, we quantify over sequences in the metalanguage, and hence the sequences may be related to the variables in our alleged semantic theory. However, according to Tarski, sequences are relations and relations never belong to the same semantic categories as their terms, and thus (the signs for) sequences, or variables ranging over them, necessarily belong to another semantic category than the signs we may substitute into their argument places.[20] In particular, f belongs to a different category than f_x. It follows that putting f for f_x in Wf_xf_x would violate the grammar of the metalanguage used. Thus, allowing sequences to be possible values of the first-order variables of an object language leads to problems of formulating an adequate truth definition for that object language.

In a little more relaxed language we may express the situation as follows. Since a sequence is a relation and relations never belong to the same semantic category as any of its arguments, a sequence assigning objects to

[20] See CTFL p. 219.

variables of some semantic category will never be one among the objects being assigned to variables of the category in question. It follows that quantification is never absolute.

3.2 Two alternatives

We just saw that, following CTFL, we never quantify in an object language over the sequences that assign values to variables of that object language. But those sequences are necessarily quantified over in the metalanguage. Hence, from the perspective of the metalanguage there is always something that is not within the range of the quantifiers of the object language and, consequently, they do not range over absolutely everything.

Tarski's adoption of the theory of semantic categories as a framework for any truth definition in CTFL is crucial for this result. But the theory of semantic categories is not the only possible framework. Indeed, Tarski himself hesitates concerning the theory of semantic categories in a postscript added to the German translation of CTFL:

> In writing the present article I had in mind only formalized languages possessing a structure which is in harmony with the theory of semantical categories and especially with its basic principles. [...] Today I can no longer defend decisively the view I then took of this question. In connection with this it now seems to me interesting and important to inquire what the consequences would be for the basic problems of the present work if we included in the field under consideration formalized languages for which the fundamental principles of the theory of semantical categories no longer hold. (CTFL, p. 268)

Tarski's main problem consists in giving a definition of truth for languages containing variables of arbitrarily high orders and, in particular, the treatment of sequences satisfying sentential functions in that context. However, the partitioning of variables into different categories remains an integrated part of the general framework of the postscript, which makes this line of development less attractive in the course of providing a semantics for absolute quantification.

The problem with absolute quantification and semantic categories consists in the partitioning of the expressions in general, and variables in particular, into mutually exclusive classes, and a corresponding partitioning of the world. One alternative to this framework takes seriously Tarski's hesitation of partitioning the objects alongside the expressions of language, which is articulated in the following passage:

> We sometimes use the term 'semantical category' in a derivative sense, by applying it, not to the expressions of the language, but to the objects which they denote. Such 'hypostatizations' are not quite correct from a logical standpoint, but they simplify the formulation of many ideas. (CTFL, p. 219.)

Tarski's use of 'denote' here is somewhat peculiar. For the expressions of a formalised language are variables, which do not denote anything, and constants, which, although they may be taken to denote, need not be thus understood. The general idea is nevertheless clear: the structure of language need not be the structure of reality. Alternatively put: grammaticality is one thing, ontology is possibly another.

In accordance with this observation one may argue that quantification over sequences, e.g., in the definition of truth for first-order object languages, brings about no new ontology, but rather new ideology. That is to say that it enhances the expressive powers without imposing new ontological claims.

There are to my knowledge no arguments along these lines for the theory of semantic categories, but there are attempts to interpret higher-order languages accordingly. Perhaps the most well-known is George Boolos's plural interpretation of monadic second-order logic.[21] We return to these kinds of theories in Chapter 5 and leave the matter for now.

If we confine ourselves to one-sorted formalised object languages it may seem unnecessary to use higher-order resources in the metalanguages. After all, what is needed is a theory of sequences to enable the use of variables in the object languages and such a theory of sequences may be developed

[21] See Boolos (1984, 1985).

within standard set theory, ZFC, or perhaps ZFCU, if urelements need to be accounted for.

The definition of satisfaction may be used *ad verbum* with the proviso that the object language considered is one-sorted and that a sequence is understood to be a set of ordered pairs of the form $\langle a, v\rangle$, were a is an object and v a variable, such that, for any given object a, we may have several pairs $\langle a, x\rangle, \langle a, y\rangle, \ldots$, in the set, but for each variable v there is one, and only one, object a such that $\langle a, v\rangle$ belongs to the sequence.

Furthermore, if no external restriction is imposed, there is no principled reason for not allowing sets in general, and sequences in particular, in the ordered pairs of a sequence. That is to say, if s is a sequence for some object language, a sequence s' such that $\langle s, y\rangle \in s'$, where y is a variable of the object language, is perfectly in order.

The set-theoretic approach differs in a fundamental way from the idea of using a higher-order language for the theory of sequences needed. For if a second-order language is used to provide sequences for some first-order object language, then, according to the above, we either impose some new irreducible ontology by quantifying over sequences, or such quantification is only illusory over entities different from the first-order entities since higher-order quantification brings new ideology rather than new ontology. The set-theoretical approach, on the other hand, reckons sequences as particular sets of ordered pairs alongside any other entity which may be a value of the variables of the object language.

In standard model-theoretic semantics the quantifier is always restricted to some set and, since there is no universal set in standard set theory, quantification over absolutely everything cannot be adequately represented. Tarski's approach in CTFL, on the other hand, need not involve any explicit restriction of the quantifiers in the metalanguage. A quantified sentence $\forall x\varphi$ is true if satisfied by some sequence s, i.e., if all sequences that differ from s for at most s_x, satisfy φ. If quantification is to be over absolutely everything, then absolutely everything has to be related to x by some sequence which is just like s save for the pair having x as its second component.

This natural suggestion comes rather close to the view of what Williamson (2003) calls the naive theorist of absolute generality. The naive theorist is an imagined defender of absolute quantification that Williamson uses to motivate an argument to the effect that absolute quantification is paradoxical. This argument has been widely discussed in the literature on absolute quantification and we devote the next chapter to it.

As we see in the next section there are strong reasons to shun the suggestion of a set-theoretic definition even without Williamson's more subtle argument. At least this is so if we use a set theory with a well-founded membership relation.

3.3 Logical consequence

Once the concept of truth has been defined it is natural to turn to the concept of *logical consequence*. Truth and logical consequence are intertwined in fundamental ways. For one thing, a characteristic property of logical consequence is truth preservation: all logical consequences of a set Γ of sentences are true provided each sentence in Γ is true. We also say that if a sentence φ is a logical consequence a set Γ of sentences then the truth of φ is necessitated by the truth of the sentences in Γ. Thus, a semantic theory would be severely hampered if it somehow made the concepts of truth and logical consequence jointly inconsistent. We thus formulate the following desiderata for any theory of semantics:

Desiderata for any semantic theory A semantic theory should allow for the definition of fundamental semantic concepts, in particular *truth* and *logical consequence*, in a consistent way.

The task of providing a semantics of absolute quantification thus turns into a problem of providing a semantic theory, which ought to include a truth definition and a definition of logical consequence, for such quantification.

When 'consequence' is used in CTFL, e.g., in convention T, it denotes the concept of *derivability*, or *being derivable from*, rather than the semantic concept of logical consequence. The latter concept is, in fact, not even defined in CTFL. Tarski notes in foresight, however, that the reduction of it to that of derivability may be unsatisfying:

> The reduction of the concept of consequence to concepts belonging to the morphology of language is a result of the deductive method in its latest stages of development. When we in everyday language say that a sentence follows from other sentences we no doubt mean something quite different from the existence of certain structural relations between these sentences. In the light of the latest results of Gödel it seems doubtful whether this reduction has been effected without reminder. (CTFL, p. 252n.)

In Tarski (1936b) the concept of logical consequence is defined within the "scientific semantics" and the notion of semantics, in turn, is explained in his contemporary (1936a):

> We shall understand by semantics the totality of considerations concerning those concepts which, roughly speaking, express certain connexions between the expressions of a language and the objects and states of affairs referred to by these expressions. As typical examples of semantical concepts we may mention the concepts of *denotation*, *satisfaction*, and *definition*, [...]. (Tarski, 1936a, p. 401, italics in the original.)

Tarski further notices that truth belongs to semantics according to this explanation. That is, it is a semantic concept since it "express[es] certain connexions between expressions of a language," in this case sentences, "and states of affairs."

Logical consequence is not mentioned among the examples of typical semantic concepts in the quoted passage and, in fact, given the above explanation one may say that, in a sense, it fails to be a semantic concept. Having explained that in order for a sentence φ to be a logical consequence from a collection of sentences Γ, if the sentences in Γ are true, then φ ought to be true, Tarski continues:[22]

> Moreover, since we are concerned here with the concept of logical, i.e. *formal*, consequence and thus with a relation which

[22] Tarski uses X and K instead of φ and Γ.

> is to be uniquely determined by the form of the sentences between which it holds, this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of the objects to which the sentence φ or sentences of the class Γ refer. (Tarski, 1936b, pp. 414–415.)

Thus, contrary to being a semantic concept in the above sense, i.e. by expressing relations, or connexions, between the expressions of some language and the objects referred to by those expressions, the concept of logical consequence is, according to Tarski (1936b), completely independent from those objects.

Tarski does not argue in (1936b) that the concept of logical consequence is a semantic concept, but shows how it can be *defined* within semantic theories. The definition uses the concept of satisfaction from CTFL and though it comes close to the modern model-theoretic definition it is not quite the same. One difference consists in Tarski's working with formalised, rather than formal, languages and, thus, that no interpretation, or reinterpretation, of the non-logical vocabulary is accounted for in the semantics. Hence the modern definition of logical consequence as preservation of truth under reinterpretations of the non-logical vocabulary is a non-starter. Instead the idea is to construct from the sentences in Γ a set Γ' of sentential functions by replacing each non-logical symbol with an appropriate variable, the same symbols being replaced by the same variables. A sequence that satisfies all sentences in Γ' is then said to be a model of Γ. φ is said to be a logical consequence of Γ only if every model of Γ is also a model of φ.[23]

Just as in CTFL, Tarski employs the theory of semantic categories in Tarski (1936b) and the variables are restricted accordingly. Thus, though no domain of quantification is imposed in the definition of logical consequence, just as it is not imposed in the definition of truth in CTFL, the implicit restriction of the variables to types, orders or semantic categories, blocks quantification from being truly absolute.

We saw in the previous section that if we drop the type-theoretical framework and employ a one-sorted set-theoretical metalanguage, it is possible,

[23] See Tarski (1936b, pp. 416–417) for details.

at least for first-order languages, to provide a definition of truth that does not employ a partitioning of the possible values of the variables. Referring to this possibility we said that a Tarski style definition of truth that is consonant with the idea of absolute quantification seems to be quite tenable.

Unfortunately, the definition of logical consequence for a language with absolute quantification in such a set-theoretical metalanguage is not as innocent as the definition of truth seems to be. In fact, it may be hard to meet the desiderata we formulated at the beginning of this section. To see this it is easier to consider the concept of logical truth rather than logical consequence. We say as usual that φ is a logical truth if it is a logical consequence from the empty collection of sentences, i.e., if every sequence is a model of it. Similarly, φ is logically false if no sequence is a model of it. Now, consider the sentence 'everything is abstract', which we may formalise into

(1) $\forall x Rx$.

(1) expresses a contingent state of affairs. Thus (1) should neither be logically false, nor logically true. It would be a logical truth if the sentential function

(2) $\forall x Xx$.

were satisfied by all sequences. Also, (1) is logically false if no sequence satisfies (2). Now, let 's_x' denote the object that the sequence s relates to the variable x. Let '$s[x/d]$' denote the sequence which is like s except that it relates d to x instead of s_x. We then have that

s is a model of (1), iff,

s satisfies (2), iff,

for everything d, $s[x/d]$ satisfies Xx, iff,

for everything d, $s[x/d]_x \in s[x/d]_X$.

But if we put $d = s_X$ and acknowledge the identities $s[x/d]_x = d$ and $s[x/d]_X = s_X$ we get as a particular case that if s is a model of (1) then $s_X \in s_X$.

Since $x \notin x$ is a theorem in standard set theory it follows that no sequence is a model of (1), making it logically false. But (1), expressing a contingent fact, is certainly not *logically* false. Hence, this way of defining logical consequence fails when we use standard set theory as metatheory for a language with absolute quantification.

A first reaction to this result might be to question the desiderata we formulated in the beginning of this section, i.e. we may question that a semantic theory for a language should enable the definition of both truth and logical consequence in a consistent way. One way of doing this would be to argue that the concept of logical consequence, though definable in a semantic theory, does not qualify as a semantic concept in the sense Tarski sets forth in the quote on page 54. However, this would obviously lead to an unwanted narrow view on semantic theorising.

A second more constructive reaction is to claim that the failure of defining logical consequence in the present setting is incidental, depending, as it does, on features of the metatheory employed. According to this reaction the merit of the argument does not lie in its showing that standard set theory is ill-suited for constructing a semantics for absolute quantification. Due to the lack of a universal set, and the common way of restricting quantification to a set in standard model-theoretic semantics, this theory is often regarded as a dead end anyway. Rather, the argument is interesting by indicating that it might be the requirement that every set ought to be well-founded that is the fundamental problem of using a set-theoretic framework for a semantics for absolute quantification. A remedy that immediately suggests itself is to make use of a non-standard set theory in which $x \notin x$ is not a theorem. We explore this line of thought in Chapter 6.

While the argument above concerns the adoption of a particular kind of set theory as metalanguage for constructing semantic theories for interpreted first-order languages, Williamson (2003) has provoked us with a more general argument to the end that regardless of the metatheory adopted, the notion of logical consequence, or logical truth, is paradoxical under the assumption that quantification is absolute. Before proceeding to develop a semantics in a theory of non-well-founded sets, we will, in the next two chapters, take a close look at Williamson's argument and some theories that have been developed in its aftermath.

4 Williamson's Argument

The old paradoxes of Cantor, Russell, Burali-Forti, etc. are traditionally considered to constitute the most severe problem for giving a formal semantics consistent with absolute quantification. Given a domain D, it cannot be all-inclusive by Cantor's Theorem, for this would violate $|D| < |\wp D|$; it cannot include a class of all classes not belonging to themselves, for that would lead to Russell's paradox; it cannot contain all ordinals as a subclass, since such a subclass would itself be an ordinal distinct from each member of the subclass.

All these examples presuppose the notion of a domain of quantification as a set, or set-like object, and hence, open up for a critique along the lines of Cartwright (1994). However, in his thought-provoking paper *Everything* Williamson (2003), Timothy Williamson presents an argument to the end that even if we reject domains altogether we may still find ourselves trapped in a variant of Russell's paradox of classes, assuming that quantification is over absolutely everything. Williamson posits himself as a defender of absolute quantification and, accordingly, he does not consider the argument conclusive. Rather, it is presented as a challenge for any defender of absolute quantification.

Two main reactions to Williamson's argument may be identified in the contemporary literature: either one thinks that the argument shows that absolute quantification is untenable, or that it merely shows that some underlying assumption or principle is faulty, and hence that absolute quantification is tenable once this flaw is fixed.

In this chapter we argue that both positions result from a misapprehension of the argument taking it too seriously. Accordingly, rather than speaking of Williamson's paradox, or variant of Russell's paradox, which has become the custom in the literature, we shall henceforth say Williamson's argument, or WA for short.

Section 4.1 gives an account of the argument and provides a brief discussion of the dialectical context in which it is formulated. Section 4.2 discusses some attempts to analyse the argument in terms of indefinite extensibility. In particular we look closely at a discussion in Glanzberg (2004), but we also consider a more Dummettian suggestion as well as an analysis given by Charles Parsons (2006). Section 4.3, finally, points out why WA fails to be conclusive.[1] In light of this discussion, the arguments in Section 4.2 are re-evaluated.

4.1 The argument

In this section we give an account of WA as it is presented in Williamson (2003). A less general version, primarily directed against the possibility of constructing a model-theoretic semantics for absolute quantification, is given in Rayo and Williamson (2003).[2]

To set the stage, Williamson invites us to ponder the views of a naive theorist. Absolutely general statements, he notes, expressed by means of 'everything' and 'something,' may standardly be given a first-order formalisation by means of '$\forall$' and '$\exists$'. It is generally acknowledged that the set-theoretical paradoxes show that there cannot be a universal set and, hence, that they prevent standard model theory to be an alternative for the naive theorist to interpret the quantifiers as ranging over absolutely everything. But, Williamson explains, playing the role of the naive theorist, despite the lack of a universal domain,

> [...] I can state the truth-conditions for quantified formulas, used in my way. As usual, it is done for truth under an assignment of values to variables. For any assignment A, variable x and thing d, let A[x/d] be the assignment just like A except that it assigns d to x.
>
> [$\forall$] $\forall x \alpha$ is true under A if and only if everything d is such that α is true under A[x/d].
>
> [$\exists$] $\exists x \alpha$ is true under A if and only if something d is such that α is true under A[x/d].

[1] The main result of that section appeared in Bennet and Filin Karlsson (2008).

[2] We return to this particular instance of the argument in Chapter 6.

> Naturally, 'everything' and 'something' in these clauses must be read unrestrictedly. (Williamson, 2003, p. 418)

In standard model-theoretic semantics 'assignment', often understood to abbreviate 'variable-assignment', usually denotes a function from the variables of the language interpreted to the domain of quantification of the model interpreting it. Here, things seem less clear. For, apparently, we neither have a model interpreting the language in question, nor a domain in relation to which assignments may be defined.

Williamson himself leaves no details of how to understand the naive theorist's concept of *assignment*, but one possibility seems to be to explore the idea from the previous chapter, taking assignments to be set-theoretic functions having the first-order variables as their domain. Although no assignment will have the totality of everything as its converse domain, we may still argue that [$\forall$] and [$\exists$] give the truth conditions of the quantifiers as ranging over absolutely everything. At least we may argue to that end if we are also prepared to argue that everything belongs to the range of at least one assignment. Furthermore, in the absence of interpretations, 'α' has to be an expression of some interpreted language. Otherwise it would make little sense to speak of the truth of '$\forall x \alpha$.' But this would put the naive theorist in a situation we found ourselves in the previous chapter, Section 3.3, and in a parallel way [$\forall$] and [$\exists$] would give no straightforward generalisation to a definition of logical consequence. We shall not repeat the details of that discussion here, but merely notice that it motivates an alternative understanding of the position of the naive theorist.

We need to take seriously the need to interpret the object language, and yet adhere to the fact that [$\forall$] and [$\exists$] are formulated without any reference to interpretations. One option seems to be to understand the naive theorist as using a concept of assignment according to which the non-logical constants are assigned values alongside the first-order variables. Unorthodox as it may be, merging the functions interpreting the non-logical vocabulary with standard assignments is a technically harmless modification of standard methods in model theory. In fact, Rayo and Williamson (2003) develop the traditional domain-based model-theoretic semantics with such a conception of assignments.

However, the naive theorist is not doing standard model theory and, in the absence of domains, merging the functions interpreting the non-logical vocabulary with standard assignments is far from a harmless modification. Actually, we once more find ourselves in a situation that parallels the one we found in Section 3.3. To see this, consider

$\forall x P x$

This is a consistent formula of first-order predicate logic. Yet there cannot be an assignment A under which it is true. At least this is so if we want A to be a set-theoretic function of some standard set theory, e.g., ZFC. For if A is such an assignment we would get

A$\models \forall x P x$, if and only if,

for everything d, A$[x/\text{d}] \models Px$, if and only if,

for everything d, d $\in$ A(P), only if,

A$(P) \in$ A(P),

which makes $\in$ ill-founded. Hence, $\forall x P x$ would be logically false given a standard set-theoretic understanding of assignments as functions from the variables and non-logical vocabulary. This is a rather startling consequence.

The above discussion shows that Williamson's account of the naive theorist's understanding of the concept of assignment, and hence the clauses [$\forall$] and [$\exists$], is too brief to be completely satisfactory. It is not so much the general way of formulating the naive theorist's position that is problematic, but the lack of ways of explicating it in such a way that it would constitute a position that needs an argument like WA in order to be refuted. As long as such an explication is lacking, additional arguments against the naive theorist's position might seem redundant.

Yet WA is stated in connection with the naive theorist's alleged considerations surrounding the Tarskian definition of logical consequence and the quantification over interpretations it involves:

> Sooner or later the naive theorist will want to generalize over all (legitimate) interpretations of various forms in the language.

> For example, the inference from $\forall x \mathrm{P}x$ and $\forall x(\mathrm{P}x \supset \mathrm{Q}x)$ to $\forall x \mathrm{Q}x$ is truth-preserving however one interprets the predicate letters P and Q. Such generalizations are the basis of Tarski's account of logical consequence [Tarski (1936b)] and its model-theoretic descendants. [...] The naive theorist wants to make such generalizations when $\forall x$ is read as unrestricted. In principle, when we apply the definition of logical consequence, it must be possible to interpret a predicate letter according to any contentful predicate, since otherwise we are not generalizing over all the contentful arguments of the right form. Thus, whatever contentful predicate we substitute for 'F', some legitimate interpretation (say, I(F)) interprets the predicate letter P accordingly:
>
> (1) For everything o, I(F) is an interpretation under which P applies to o if and only if o Fs. (Williamson, 2003, p. 426)

It seems clear that (1) is a schema of the form $\forall o\,\Phi(F, o)$, where 'F' is a placeholder for *contentful predicates*.

The somewhat unexpected talk of interpretations in this passage (the word 'interpretation' makes its first appearance in the paper with this quote) may be taken to indicate that Williamson thinks of interpretations and assignment as interchangeable as was suggested above. It may also be taken to show that WA may be understood quite independently from the context of the naive theorist. Either way, WA is thought to gain generality if the notion of interpretation is not further specified. In particular, interpretations need not be set-theoretic entities.

Also, Williamson's allusion to Tarski (1936b) might need a word of clarification. We saw in Section 3.3 that Tarski came close to the now standard model-theoretic definition of logical consequence. In Tarski's definition we quantify over sequences of objects which may, or may not, satisfy sentential functions resulting from other sentential functions by uniformly replacing the non-logical vocabulary by variables of appropriate categories. Since we do not quantify over interpretations in (1), which would be the equivalents of Tarski's sequences, but merely over contentful predicates, which are most naturally understood as syntactic entities of the metalanguage, one might suspect that Williamson's account is rather different from Tarski's.

However, before arriving at the final definition of logical consequence in (1936b) Tarski tentatively suggests a definition based on uniform substitution in the object language rather than reinterpreting it. He rejects it as a definition since it would make the relation of logical consequence unduly dependent on the language under consideration, but recognises it as a necessary, but not sufficient, condition on logical consequence. Though Tarski does not consider truth under uniform substitution in the metalanguage, we may still understand (1) in a similar way as the preservation of truth under substitutions in the object language, i.e. as a necessary, but not sufficient, condition. For whatever contentful predicates are available in the metalanguage, a possible requirement on any semantics is that it contains interpretations interpreting 'P' according to those contentful predicates. That seems to be what (1) aims at saying.

The next step of the argument consists of a definition of a contentful predicate which, under substitution for 'F', yields a contradictory instance of (1):

> (2) For everything o, o Rs if and only if o is not an interpretation under which P applies to o.
>
> The naive theorist is committed to treating 'R' as a contentful predicate, since it is well-formed out of materials entirely drawn from the naive theory itself. (Williamson, 2003, p. 426)

Here 'the naive theory' seems to embrace (1), so that the commitment to accept R as a contentful predicate rests upon the fact that no new concepts are introduced in its definition.

Next, substitute 'R' for 'F' in (1) and apply (2) to get:

> (3) For everything o, I(R) is an interpretation under which P applies to o if and only if o is not an interpretation under which P applies to o. (Ibid.)

Instantiation gives:

> (4) I(R) is an interpretation under which P applies to I(R) if and only if I(R) is not an interpretation under which P applies to I(R). (Ibid.)

This is a contradiction.

The reasoning leading to the contradiction seems rather straightforward. Williamson assumes that we may interpret the quantifiers of some (first-order) language to range over absolutely everything. Next he notices that, given some predicate letter 'P', we need to be able to interpret it as any contentful predicate. Otherwise, he claims, we cannot get the definition of logical consequence right. Thus, for each contentful predicate there is an interpretation interpreting 'P' accordingly. In particular, this holds for the contentful predicate R that Williamson defines and its corresponding interpretation I(R). From this the contradiction follows by plain logic.

A first analysis that Williamson gives runs as follows: Since we do generalise over interpretations in the metalanguage, the (first-order) variables of the object language have to range over interpretations if quantification is absolute—absolute quantification entails quantification over interpretations. This is used in the step from (3) to (4). Hence, it may be argued, the variables of the object language may not, on pain of contradiction, range over all interpretations. It follows that quantification in the object language is not over absolutely everything.

4.2 Indefinite extensibility and WA

We said that WA is generally considered a variant of, or of the same kind as, Russell's paradox of classes.[3] Russell argued, in (1907), that his paradox may be understood to show the existence of self-reproducing properties.

[3] Williamson himself says that "the argument is obviously a variant of Russell's Paradox" (Williamson, 2003, p. 426), and uses phrases such as 'the Russellian paradox' and 'the version of Russell's paradox' when referring to it. Michael Glanzberg (2004, p. 552) too calls it "a version of Russell's paradox". There is even a sub-entry to 'Paradoxes' in Rayo and Uzquiano (2006) where it is listed as 'Williamson's variant of Russell's'. See however

These properties are close to Dummett's indefinitely extensible concepts. According to Dummett, it was the failure of recognising the existence of such concepts together with the assumption of a domain of quantification containing the extension of every definable concept, that made Frege adopt the inconsistent system of *Grundgesetze*.[4]

WA's resemblance to Russell's paradox has been taken to indicate that it too may be analysed in terms of indefinitely extensible concepts. The general idea is that it shows that some concept, which ought to be definite if absolute quantification is tenable, is in fact indefinitely extensible. Thus analyses of this kind tend to accept the sceptical conclusion from WA. They may differ, however, in which concept is understood to be indefinitely extensible.

4.2.1 Indefinite extensibility and the logical concept of object

One example of an analysis of WA in terms of indefinite extensibility is given by Michael Glanzberg in his *Quantification and Realism* (2004). Actually the analysis is embedded in a larger argument against absolute generality. Very briefly, and in rough terms, this argument may be put as follows: In order to generalise over some things, or alternatively, in order to interpret an utterance involving a quantifier as ranging over some things, we need to have a determinate conception of the things in question. In the case of absolute generality we would need a determinate conception of all things. But WA shows, Glanzberg argues, that there is no determinate conception of all things. Hence, we cannot generalise over absolutely everything.

This looks like a familiar line of argument: substitute 'set' for 'determinate conception', and 'Russell's paradox' for 'WA', and we get a familiar formulation of a traditional argument against absolute quantification: In order to generalise over some things, or alternatively, in order to interpret an utterance involving a quantifier as ranging over some things, we need to have a set of the things in question. In the case of absolute generality we

Bennet and Filin Karlsson (2008), and Section 4.3, where it is shown that, even if WA and Russell's Paradox are similar they differ in aspects that are crucial for the discussion on absolute quantification.

[4] See Chapter 2.

would need a set of all things. But Russell's paradox implies that there is no set of all things. Hence, we cannot generalise over absolutely everything.

The model-theoretic argument is robust in the sense that, if the premises are accepted, i.e., if there ought to be a set the members of which are precisely the things we generalise over, and if there cannot be a universal set, then the conclusion seems inevitable. In this section we discuss Glanzberg's analysis of WA in order to judge whether it gives rise to an argument that is as robust as the argument from Russell's paradox. Such a reading also has the merits of clarifying the precise role of WA in this argument.

Three truisms of meaning

Glanzberg invites us to consider three principles, or truisms, of meaning which he takes to be almost self-evident. The first principle tells us that "*utterances only have the meanings they do because they are interpreted as having them.*" (Glanzberg, 2004, p. 543, Glanzberg's italics) Hence, utterances are meaningful only if they are interpreted.

Glanzberg intends to be neutral on the nature of meanings, i.e. on what it is that an utterance has once it is interpreted. Thus, without taking a stand on that issue, the second truism tells us that, whatever they are, "*meaning determines truth conditions.*" (p. 544, Glanzberg's italics). A consequence of the first and second truism is that interpretations determines the truth conditions for utterances.

The third, and final, truism tells us what is needed of an interpretation to fix the truth conditions for utterances involving quantification: "*interpretation must provide a domain of quantification.*" Glanzberg adds that "[he does] *not yet claim the domain cannot be absolutely everything*, only that some domain must be provided." (Ibid).

The third truism is, in contrast to the first and second, somewhat harder to accept as a truism, i.e., as almost self-evident. By requiring a domain of quantification it would seem that Glanzberg blatantly adheres to Cartwright's All-in-One Principle: "the values of the variables must be in, or belong to, some one thing." (Cartwright, 1994, p. 7) We saw in Section 2.2.3 that, according to Cartwright, the All-in-One Principle is, not only false, but that it also figures as a fundamental assumption in Dummett's argument from indefinite extensibility to the impossibility of absolute quantification. Glanzberg's argument is, just like Dummett's, an argument against absolute quantification from indefinite extensibility and it appears that by using the third truism it too may become vulnerable to Cartwright's critique.

Specifications of domains and avoiding the All-in-One Principle

Interestingly, Glanzberg argues that stating the third truism does not necessarily entail adherence to the All-in-One Principle as formulated and criticised by Cartwright.[5] To see this we need to have a clear understanding of what it means for an interpretation to provide a domain. Unfortunately, Glanzberg's explanation of this notion is not completely clear. For instance, he sometimes speaks as if domains are provided by *speakers*, even though it seems perfectly clear from the third truism that it is *interpretations* that provide domains.[6] Here is a telling passage:[7]

> Interpretation, as I stressed, is something done by speakers. Thus, the question whether absolutely unrestricted quantification can be accomplished comes down to the question of what is required for a *speaker* to provide a domain, and whether it can be done for a domain of 'absolutely everything'. The point of the argument so far has been to show that this is the question we really need to answer.
>
> […]
>
> So, let us try to answer it. Is it possible for speakers to specify the domain of 'absolutely everything'? (Glanzberg, 2004, p. 545, Glanzberg's italics.)

The first sentence of the quote reminds us that interpretation is something that speakers perform, it is an activity. It would be odd to identify this *act*

[5]See Glanzberg (2004), p. 555, n. 8.

[6]It would be more natural to replace 'speaker' by 'hearer' in this context, but we follow Glanzberg in ignoring the hearer/speaker distinction for reasons of simplicity. Nothing important hinges on this distinction in the present context.

[7]The left out part is a section heading: *"1.3. Domains for unrestricted Quantifiers"*.

of interpreting with the *speaker* performing it. For instance, it is natural to presume that two distinct speakers may perform the same act of interpretation. But, in light of the third truism, Glanzberg may be taken to suggest this identity when asking "what is required for a *speaker* to provide a domain".[8] A more benevolent understanding seems possible however: a speaker may be said to provide a (particular) domain by performing one of the acts of interpreting that, in turn, provides the particular domain in question. Speakers may in that way be said to provide domains in a derivative sense.

This view, that speakers provide domains of quantification by means of performing acts of interpretation, may also explain why Glanzberg puts requirements on the speaker rather than the act of interpretation when domains are provided. A speaker may need to possess certain faculties in order to be capable of performing particular acts of interpretation. The question whether absolute quantification is possible then comes down to the question what is required of a speaker in order for him to perform an act of interpretation that provides the domain of 'absolutely everything'. Glanzberg answers in the last sentence of the quote that a speaker must *specify* the domain in order to provide it. That is, if speakers provide domains in a derivative sense by performing acts of interpretation, the act of interpreting a sentence that involves quantification must somehow contain a specification of a domain of quantification.

We said that Glanzberg, despite requiring specifications of domains of quantification, does not regard himself as committed to the All-in-One Principle, i.e., to the existence of some one thing that collects together the things quantified over. The situation seems to be that while the arguments against absolute quantification from the set-theoretic paradoxes rely on the notion of a domain of quantification as an object, WA makes this assumption redundant. There is simply no talk of domains in WA. Thus, equipped with WA, Glanzberg need not postulate domains as objects in order to derive a contradiction from the assumption that quantification is absolute. Furthermore, in light of Cartwright's critique, he is well motivated to resist that specifications of domains automatically give rise to domains as objects.

[8] Note that the emphasis is Glanzberg's.

There are certainly some delicate philosophical subtleties going on here, for what do we specify if not *the range* for the quantifiers? We shall not review Glanzberg's arguments regarding this however, but, for the sake of argument, we simply accept that the requirement of specifications to provide domains of quantification does not entail commitment to the All-in-One Principle.[9]

Now, if a speaker interprets a quantifier in an utterance as being absolute, the third truism says that the interpretation must *provide* a domain of quantification that embraces absolutely everything. Furthermore, domains are provided by being specified according to Glanzberg. Precisely how the domain of absolutely everything may be specified is, however, a matter of some delicacy and, in fact, the overall conclusion of Glanzberg's argument is that such a specification cannot be provided.

Two observations

We make two observations before proceeding to the argument that there cannot be a specification of the domain of absolutely everything. First we note that while WA is clearly an argument that is primarily directed against the possibility of giving an explicit semantics for absolute quantification, Glanzberg seems to be involved in the bigger question if it is possible to *talk* about absolutely everything.[10] These questions are related, but certainly not the same. For one thing, to construct a semantics for absolute quantification we presumably need to be able to talk about absolutely everything, but in order to talk about absolutely everything we need not an articulated semantic theory for absolute quantification. Thus, when using WA to show that absolute quantification is contradictory Glanzberg seems to restrict himself to the question if it is possible to *articulate* an explicit semantics for absolute quantification.

Second we note that, while Glanzberg sometimes considers what is required of a *speaker* to specify different domains, it is clear from other passages of his paper that the real requirements are put on the specifications. When discussing the part of the paper where WA comes into play, we may

[9] See (Glanzberg, 2004, note 8, p. 555–556).

[10] See, for instance, the beginning of the second part of Glanzberg (2004).

thus benefit from skipping the talk of speakers performing acts of interpretations and specifying domains; instead we speak of interpretations and specifications directly.

Determinate specifications

Now, Glanzberg's tactics when arguing for the impossibility of a semantics for absolute quantification is to show that there cannot be a specification of a domain of quantification embracing absolutely everything. This takes us back to the question on precisely how domains are provided. However, this time domains are not understood as being provided in, or by, an act of interpretation, but rather by some specification in a semantic theory.

Starting out in known territory, Glanzberg identifies two fundamental properties that specifications of domains that result from applying predicates to some background domain ought to have. First, a specification needs to be *sharp*. In the case of specification from a background domain this means that, for every object in the background domain it ought to be determinate whether or not it meets the specification and belongs to the specified domain. That is, for any object, it is determinate whether or not the predicate used to specify the domain is true of that object. The second property is that a specification of a domain ought to be *exhaustive*. The idea, it seems, is to rule out specifications that are, in Russell's sense, self-reproductive:

> ...there are what we may call *self-reproductive* processes and classes. That is, there are some properties such that, given any class of terms all having such a property, we can always define a new term also having the property in question. Hence we can never collect *all* the terms having the said property into a whole; because, whenever we hope that we have them all, the collection which we have immediately proceeds to generate a new term also having the same property. (Russell, 1907)

Thus, given an exhaustive specification Σ of a domain D_Σ, the assumption that this domain contains all objects meeting the specification Σ must not provide means to identify objects that cannot be among objects in D_Σ, but nevertheless satisfy the specification Σ.[11]

A specification that is both sharp and exhaustive, Glanzberg calls *determinate*. To show that there cannot be a consistent semantics for absolute quantification, Glanzberg now proceeds by showing that there cannot be a determinate specification of a domain of absolutely everything. Of course, we can never specify a domain of absolutely everything by means of applying predicates to some background domain. But, while the notion of determinateness is given in relation to such specifications, Glanzberg explains that it is also reasonable to require determinateness from specifications not using background domains, e.g., any purported domain of absolutely everything.

The impossibility of a determinate specification of the domain of absolutely everything

To get his argument off the ground, Glanzberg tentatively suggests that the *logical concept of object*, as introduced by Charles Parsons[12], may be used to give a determinate specification of the all-inclusive domain. As Glanzberg points out, there can be no doubt that the ordinary concept of object is much too vague to yield a sharp specification. The logical concept of object, however, uses the idea that there is a determinate specification of the possible referents of the singular terms. Then the objects, and hence the domain of objects, may be sharply specified: "the objects are all and only the potential referents of singular terms" (Glanzberg, 2004, p. 550). Needless to say, such a specification may be challenged on many points, but it is part of the dialectic of Glanzberg's argument to treat it as "the best case" for the defender of absolute quantification, and thus accept it to be at least sharp.

Glanzberg now uses WA to show that the domain specified by means of the logical concept of object is sharp but non-exhaustive:

[11] As we have said, Russell's notion of self-reproductiveness is a forerunner to Dummett's notion of indefinite extensibility. We shall have more to say on how these two concepts relate in the present context in Section 4.2.2.

[12] This concept is recurring in Parsons works, but see his (1982) for a particularly rewarding presentation.

> [WA] shows that if we have a determinate specification of a plausible candidate domain, we can use it to find another object, which cannot on pain of contradiction be in the domain specified. Thus, the paradox shows that plausible determinately specified domains cannot be absolutely everything. (Glanzberg, 2004, p. 552)

The object referred to as falling outside the domain of quantification is I(R). To secure that this is a proper singular term Glanzberg relies on the *method of nominalisation*. According to this method, rather than treating predicates themselves as referring to objects, we may transcribe the context in which they occur so that they become nominalised, i.e., names on par with singular terms.[13] Glanzberg explains that

> [w]e do not really need the full nominalizing power of English to argue against absolutely unrestricted quantification. What we need is to be able to nominalize the process of interpretation, as the more general version of the Russell argument [i.e., WA] showed. This gives us objects which cannot fall within the domain of the unrestricted quantifiers of the interpretation, even if the domain endeavored to be maximal. (Glanzberg, 2004, p. 555)

Thus, when applying WA to a purported interpretation of absolute quantification Glanzberg shows that by nominalising the interpretation, or process of interpreting, we get a potential referent of a singular term, and hence a quantifiable object in the logical sense of that concept, that cannot lie within the domain provided by the interpretation in question.

Later, reference is made to Dummett's notion of indefinitely extensible concepts:

> The logical notion of object thus exhibits what Dummett [...] has called indefinite extensibility. Insofar as the logical notion of object can be used to produce a determinate specification of a domain of quantification, it produces not just one, but an

[13] See Parsons (1982) for a discussion on the method of nominalisation.

indefinitely increasing sequence of them. (Glanzberg, 2004, p. 557)

Formalising the argument

To make this a little more precise it is instructive to spell out the details of the argument leading to this conclusion by formalising WA.

Assume thus a first-order language $\mathscr{L}$ which contains a one-place predicate symbol 'P'. Let 'P' be a name of 'P', i.e. 'P' is a term in the metalanguage $\mathscr{L}^m$ of $\mathscr{L}$. Furthermore, we need to be able to say, in the metalanguage, that x is an interpretation, which is accomplished by '$INT(x)$', and that x applies to y under z, for which we use '$z \models x[y]$'. Also, let 'I' denote the operator taking contentful predicates to interpretations. Syntactically, 'I' is a term forming operator.

In formalising (1) we have that, for each contentful predicate F of $\mathscr{L}^m$,[14]

(5) $\forall x(INT(I(F)) \wedge (I(F) \models P[x]) \leftrightarrow F(x))$

The definition of R becomes:

(6) $\forall x(R(x) \leftrightarrow \neg(INT(x) \wedge (x \models P[x])))$

Applying I to R gives the interpretation $I(R)$ of $\mathscr{L}$. By Glanzberg's third truism, it must provide a domain of quantification, D, and by assumption we take D to be the domain specified by the logical concept of object. Furthermore, $I(R)$ is an object in the logical sense of this concept, being referred to by means of the singular term '$I(R)$', and hence, it must be a possible value of the variables of $\mathscr{L}$.

Substituting R for F in (5) gives:

(7) $\forall x(INT(I(R)) \wedge (I(R) \models P[x]) \leftrightarrow \neg(INT(x) \wedge (x \models P[x])))$

[14] Indeed, this is not the only way of formalising (1), but it brings to the surface the logical structure needed in Glanzberg's argument. Alternatives will be suggested below.

which, since $I(R)$ belongs to D, may be instantiated by $I(R)$.[15] But that gives the contradiction. Hence, we may conclude, $I(R)$ does not belong to D after all and D does not consist of absolutely everything.

This line of reasoning seems to generalise. Let D be *any* purported domain of absolutely everything. Define R over D as above. Then, as above, we may show that $I(R) \notin D$. Even if D is enlarged by adding $I(R)$ to get D', we may define R' in a similar way as R was defined, and go on to show that $I(R') \notin D'$. Thus we get a sequence of larger and larger domains specified by means of the logical concept of object, and in that sense this concept exhibits the property of being indefinitely extensible.

The logical concept of object constituted the most promising specification of an all-inclusive domain of quantification. If it is indefinitely extensible, it may not specify an exhaustive domain of quantification. "Hence, no determinate [i.e. sharp and exhaustive] specification gives 'absolutely everything' " (Glanzberg, 2004, p. 557). Since a determinate domain of quantification is needed to interpret quantified sentences, by the third truism, there will be no interpretation interpreting the quantifiers as ranging over absolutely everything.

4.2.2 Alternative indefinitely extensible concepts

Glanzberg's use of 'indefinite extensible' in arguing against absolute quantification, is slightly different from how this term is used by Dummett, e.g., in his (1991). Consider what Dummett says about the indefinite extensibility of the concept of ordinal number:

> What the paradoxes revealed was not the existence of concepts with inconsistent extensions, but of what may be called indefinitely extensible concepts. The concept of an ordinal number is a prototypical example. The Burali-Forti paradox ensures that no definite totality comprises everything intuitively recognisable as an ordinal number, where a definite totality is one quantification over which always yields a statement determinately true or false. For a totality to be definite in this

[15] If $I(R) \notin D$, then $I(R) \models P[I(R)]$ would presumably not make sense.

> sense, we must have a clear grasp of what it comprises: but, if we have a clear grasp of any totality of ordinals, we thereby have a conception of what is intuitively an ordinal number greater than any member of that totality. Any definite totality of ordinals must therefore be so circumscribed as to forswear comprehensiveness, renouncing any claim to cover all that we might intuitively recognise as being an ordinal. (Dummett, 1991, p. 316)

Thus, having a determinate domain D of quantification over which the concept *Ord* of ordinal number is defined, Burali-Forti's paradox ensures that there is an ordinal number outside of D. Thus the definition of *Ord* over D immediately gives rise to a new domain of quantification D' consisting of D and the new ordinal. *Ord* will naturally have a new extension over D', and will again give rise to a new ordinal. An endless sequence of more and more inclusive domains $D, D', D'', \ldots$, is accomplished. Likewise, Russell's paradox may be taken to show that there is no domain of quantification containing all classes.

According to Dummett, Frege's mistake consists in supposing that there is a domain containing the extension of every *concept* defined over it.[16] In Glanzberg's analysis the "mistake" leading to the contradiction in WA consists rather in assuming that there is a domain containing every *object* defined over it.

In comparison, it is quite clear that Glanzberg's line of argument is afflicted with a certain kind of weakness that an analysis of WA along the lines of Dummett could avoid. For one thing, by ruling out the particular specification of a domain by means of the logical concept of object, only one suggested domain of quantification has been rejected. There may be other specifications that actually do the job. In fact, the domain of quantification as specified by means of the logical concept of object seems to suit Williamson's argument a little too well. A necessary condition for the contradiction, as it is derived above, is that '$I(R)$' is a singular referring term, which it is thanks to the method of nominalisation. Then, according to the logical conception of objects, whatever '$I(R)$' refers to, it has to be counted

[16] Dummett (1991), p. 317

as an object. Given this, it is possible to instantiate (7) with $I(R)$.

Moreover, Williamson's argument can be stated without the assumption that '$I(R)$' is a singular referring term. Consider (1) again. Roughly, this principle aims at saying that whatever contentful predicate F is, there is some interpretation y that interprets 'P' as F. There is no need to construct singular terms referring to interpretations since, by assumption, they are within the range of our quantifiers:

(8) $\exists y \forall x((INT(y) \wedge (y \models P[x])) \leftrightarrow F(x))$

With R defined as in (6) we have:

(9) $\exists y \forall x((INT(y) \wedge (y \models P[x])) \leftrightarrow \neg(INT(x) \wedge (x \models P[x])))$

A contradiction is reached in the obvious way.

How the background domain is specified, be it with the logical concept of object or in some other way, seems immaterial as long as it includes INT.[17] In particular, the specified domain need not purport to contain everything there is. Minimally, let INT be the domain of quantification for $\mathscr{L}$. Then, we may delete the occurrences of '$INT(y)$' in (8) and (6) and proceed to derive a contradiction in the same way as above. That is, rather than (8) and (9), we get,

(10) $\exists y \forall x((y \models P[x]) \leftrightarrow F(x))$

and

(11) $\exists y \forall x(y \models P[x] \leftrightarrow (x \not\models P[x]))$

where the definition of R is

(12) $\forall x(R(x) \leftrightarrow (x \not\models P[x]))$

Hence, to get a contradiction, there is no need to assume, with Glan-

[17] If the domain does not include INT the argument simply shows that $I(R)$ is outside the domain.

zberg, an all-inclusive domain of objects that is sharply specified. A sharp specification of INT is quite sufficient, and such a specification is assumed anyway since 'INT' is a monadic predicate in $\mathscr{L}^m$.

It is tempting to take this to indicate that, rather than the logical concept of object, it is the concept of *interpretation* that WA shows to be indefinitely extensible. For what it's worth, this would lead us to an analysis more similar to Dummett's analysis of Burali-Forti's paradox.

4.2.3 Parsons's diagnosis

Parsons (2006) suggests yet another alternative as to which concept is shown to be indefinitely extensible by WA, namely the concept of being *true under an interpretation*. Parsons's analysis rests upon a slightly different terminology than the original formulation of WA, but $\mathscr{L}^m$ is still suitable for formalising it. The meta-logical principle corresponding to (1) is rendered:[18]

(13) P is true of x according to $I(F) \leftrightarrow Fx$.

In the language of $\mathscr{L}^m$ this becomes:

(14) $\forall x((I(F) \models P[x]) \leftrightarrow F(x))$.

The corresponding definition of R becomes:

(15) $\forall x(R(x) \leftrightarrow (x \not\models P[x]))$.

A contradiction is reached in the obvious way.

We may take this to indicate that $I(R)$ isn't a possible value of x, and hence that the quantifiers do not range over all interpretations. But this need not be considered an argument for the indefinite extensibility of the concept of interpretation. One may grant the possibility of quantifying over all interpretations (of $\mathscr{L}$) if one is simultaneously prepared to restrict the range of applicability of $\models$. For once we allow quantification over all interpretations, the argument shows that there is at least one interpreta-

[18] For Parsons's account of the argument, see Parsons (2006), p. 212.

tion, $I(R)$, for which the concept of *P is true of x according to* $I(R)$ is not determinate.

This indeterminateness is intertwined with the notion of indefinite extensibility:

> The friends of absolute quantification owe us a resolution of the tension between quantification over all interpretations and the absence of a determinate notion of truth under an interpretation. One way of putting the point is that the position seems to be that 'interpretation' is not an indefinitely extensible concept while 'truth under an interpretation' is. (Parsons, 2006, p. 213.)

The indefinite extensibility of the concept of truth under an interpretation comes from the fact that, if we assume that this concept is determinate, that is, if we assume that $x \models y[z]$ is determinate, then we may define a new object, $I(R)$, in terms of $x \models y[z]$ (and 'P'), such that $I(R) \models P[x]$ is not applicable to this object. Of course, this does not prevent us from adopting a wider concept of truth under an interpretation, $x \models' y[z]$ say. But then there will be just another problematic object $I(R')$ defined from $R'(x) \leftrightarrow (x \not\models' P[x])$. And so on and so forth.

4.3 Deflating WA

In the previous sections we have seen how WA is analysed to show that some concept involved in the argument is indefinitely extensible. This, in turn, is taken to show that absolute quantification is untenable. Thus, on the surface, the role WA plays in arguments against absolute quantification from indefinite extensibility is similar to the role Russell's paradox plays in such arguments. Indeed, as we stressed, WA is itself often considered a variant of Russell's paradox. We argue in this section that the analyses from the previous sections result from a misapprehension, taking the argument too seriously in the discussion of absolute quantification. This becomes evident when we clarify just how similar WA and Russell's paradox of classes really are.

We begin this section by recalling some well known facts about Russell's paradox and comparing it to the alleged paradox of the barber. Next, we proceed to show that WA is more like the non-paradox of the barber than Russell's paradox. We close the chapter by some reflections about just what this means for the possibility of giving a semantics for absolute quantification.[19]

4.3.1 Russell's paradox

Russell's paradox is derived by considering the class r consisting of all those classes that do not belong to themselves. That is, $\forall x(x \in r \leftrightarrow x \notin x)$. By instantiation we get the absurdity $r \in r \leftrightarrow r \notin r$. Now, since the assumption of r implies a contradiction, this assumption cannot be true—there is no such r. We seem to have accomplished a *reductio ad absurdum* of the existence of a class of all classes that do not belong to themselves.

So far there is nothing paradoxical about Russell's argument. However, the paradox is apparent once it becomes clear that the existence of a class such as r is not an assumption, but a consequence of a fundamental principle, viz. the principle of comprehension. This principle, the formal rendering of which is $\exists y \forall x(x \in y \leftrightarrow \varphi(x))$ for each formula $\varphi(x)$, guarantees that there is such a class r, by letting $\varphi(x)$ be $x \notin x$.

Thus, if the principle of comprehension is regarded as a sound principle, and hence gives us a strong reason to believe in the existence of r, and the argument from the assumption of the existence of r to the absurdity $r \in r \leftrightarrow r \notin r$ is regarded as a correct *reductio*, we have a paradox.

The situation may be clarified if we consider in this connection the well known (alleged) paradox of the barber. Thus, consider a barber, living in some village, who shaves all and only those villagers who do not shave themselves. That is, we are asked to consider a barber b such that $\forall x(S(b, x) \leftrightarrow \neg S(x, x))$. Instantiating the formula thus gives the false statement $S(b, b) \leftrightarrow \neg S(b, b)$ This resembles the first part of Russell's paradox. However, since we have no independent reason to believe in the existence of the barber, the argument simply shows that there is no such barber.

Thus, Russell's *reductio* of the existence of r, i.e. the derivation of $r \in r \leftrightarrow$

[19] A large part of this section is taken from sections 4–5 of Bennet and Filin Karlsson (2008).

$r \notin r$, unlike the barber argument, becomes a paradox in the presence of an independent and apparently plausible principle from which the existence of r follows. Now, one may ask, is Williamson's argument like Russell's paradox of classes, or is it merely of the barber kind, viz. a simple *reductio*?

4.3.2 Williamson's barber

To clarify the structural similarities between Russell's paradox and Williamson's variant it is helpful to use the formalised version of the argument. Or perhaps we should use the plural and say versions. So far we have seen three different formalised versions of the principle (1) of WA. First, in the discussion of Glanzberg's analysis, we gave it a rather head on formalisation:

(5) $\forall x(INT(I(F)) \wedge (I(F) \models P[x]) \leftrightarrow F(x))$.

Here 'F' is a placeholder for syntactic entities of the metalanguage, the contentful predicates, and is thus a schema instantiated by different contentful predicates for 'F'.

In Section 4.2.2 we made the simple observation that the operation I, taking contentful predicates to interpretations is redundant. This gave us

(8) $\exists y \forall x(INT(y) \wedge (y \models P[x]) \leftrightarrow F(x))$.

Both (5) and (8) are faithful to Williamson's original formulation of the argument in the sense that the contradiction follows with R defined as in (2), which is straightforwardly formalised into

(6) $\forall x(R(x) \leftrightarrow \neg(INT(x) \wedge (x \models P[x])))$.

We saw also that Parsons' uses a somewhat different version of (1),

(16) $\forall x((I(F) \models P[x]) \leftrightarrow F(x))$,

which is just an instance of (10).

However, one might also argue that (5), or (8), does not really capture the intended meaning of (1). Instead one might want to suggest the following alternative:

(17) $\exists y(INT(y) \wedge \forall x((y \models P[x]) \leftrightarrow F(x)))$.

In each of (10), (16), and (17), a contradiction follows directly by substituting '$x \not\models P[x]$' for 'F'.

Now each of these formalised versions of (1) is structurally similar to the principle of comprehension, being of the form $\exists y \forall x(\psi(x, y) \leftrightarrow F(x))$, or of the form of an instance, $\forall x(\psi(x, I(F)) \leftrightarrow F(x))$, of it. But whereas the principle of comprehension, in the case of Russell's paradox, entails the existence of the class r, (1) does not seem to have anything to do with the existence of R in any of its versions above. We may deny its existence without giving up (1) as a true principle, while it was precisely the impossibility of making such a move in Russell's argument that made it paradoxical. In this sense Williamson's argument is more like the non-paradox of the barber, i.e. it is a *reductio* of R, or, more precisely, it is a *reductio* of R being a contentful predicate.

But, one may ask, why should we deny that R is a contentful predicate? The definition looks fairly innocent, and surely there are many things that do not satisfy $INT(x)$. But this is precisely one of the lessons of Russell's paradox: writing down a definition does not guarantee that the definiendum exists. If a paradox appears, *something* has gone wrong. The definition may be ill-formed (as Russell concluded), or it may turn out that nothing satisfies the defining condition. To show that something does satisfy it requires an independent argument, and in the case of R, none has been given.

4.4 Concluding remarks

Once Williamson's alleged paradox is shown to be of the harmless kind, the analyses in terms of indefinite extensibility that we elaborated on in Section 4.2 lose much of their force. For now WA is seen to show, not that the logical concept of object is indefinitely extensible because some object, $I(R)$, can be found outside its allegedly definite extension, but that some predicate, R, cannot be counted among the permissible substitution instances to (1). For similar reasons WA neither shows the indefinitely extensibility of INT nor the indefinite extensibility of $\models$. It simply shows that given the truth of (1) in some of its versions, there is some predicate that, despite appearances, cannot be counted among the contentful predicates.

We could, of course, argue that it is the principle expressed in (1) rather than the definition of *R* that is mistaken. Formally this would certainly be possible, but then we would have to argue for the falsity of a principle that we said expresses a necessary condition for a Tarskian definition of logical consequence. In light of that, the renunciation of *R* being a contentful predicate seems much more natural.

Linnebo (2006) also argues that the most natural way to meet WA is to reject the definition of *R*. But whereas he requires additional arguments for doing so, the moral of this chapter is that WA is in itself a sufficient argument. We do not need to develop a theory of concepts and properties that explains precisely why we are justified in rejecting the definition of *R*, although we certainly agree such a theory will be very interesting. We will have more to say about Linnebo's theory in the next chapter. The point we make here is simply that the truth of (1) rules out the definition of *R*.

5 Some Alternative Semantics

In this chapter we review three semantic theories that are consistent with absolute quantification. In addition to standard requirements it is often required that such a semantics is *strictly adequate* or a that it is a *general semantics*. Strict adequacy, as we shall see, may be understood in two ways, but the basic idea is clear enough: a semantics is strictly adequate if, for any possible way a non-logical constant may be interpreted, there is an interpretation interpreting it in that way. Loosely speaking, a semantics is a general semantics if it allows the, in some sense, intended semantics for any legitimate language.[1]

The paradoxes, WA, and the requirement of strict adequacy have all been used to argue for the necessity of typed languages in the construction of a semantics for absolute quantification. But, under the assumption that our first-order quantifiers range over absolutely everything, the question over what the higher-order quantifiers range becomes non-trivial. In this chapter we consider two theories that give different answers to this question. Thus, in Section 5.1 Williamson's (2003) suggestion that the higher-order quantifiers may be taken to range over Fregean concepts is presented and briefly discussed, and in Section 5.2, Rayo's (2006) account, suggesting that higher-order quantifiers may be understood to range over higher-order pluralities, is given.

Both Williamson's and Rayo's accounts provide a hierarchy of languages in which each language has a strictly adequate semantics, in one sense of that word, formulated in some language higher up in the hierarchy. However, there seems to be no way of providing a strictly adequate semantics for the language of type theory itself. This is due to the need for expressing facts whose formulation requires that we violate the type restrictions in one way or the other. Russell, e.g. in (1908), was well aware of this problem.

[1] See (Linnebo, 2006, p. 150).

In Section 5.3 we rehearse the main argument against type-theoretic semantics by considering how Linnebo (2006) applies these problems to the semantic theories suggested by Rayo and Williamson. We also show, in that section, by applying an argument reminiscent of Grelling's paradox, that another type of inexpressibility puts the motivation of the typed semantics in some doubt. One conclusion of this chapter is, thus, that we should not rest content with the suggested typed semantics, but that we should proceed to search for yet other semantics for absolute quantification.

To construct a semantics for absolute quantification we need a suitable theory to work in. As we have said, due to the lack of a universal set and the set-theoretic paradoxes, standard set theory, ZFC, constitutes a poor framework for a semantics adequate for absolute quantification. Indeed, the set-theoretic paradoxes have even been taken to warrant the impossibility of absolute quantification as such, i.e., not merely the impossibility of providing an adequate semantics within ZFC. Thus one natural response is to at least give up ZFC as a means to our ends, that is, if one does not already agree to give up absolute quantification as a consequence of the paradoxes in the first place. Linnebo (2006) takes a stand against such a conclusion. Rather than giving up ZFC he suggests that it should be supplemented with a theory of properties. The resulting theory should be strong enough to construct an adequate semantics and the theory of properties must avoid WA in a natural, non ad hoc way. We discuss this in Section 5.4.

5.1 Williamson's semantics

Williamson (2003, IX) argues that a semantics for first-order quantification over absolutely everything is best developed in a second-order metalanguage.[2] In particular the *truth definition* for a first-order language should be formulated in a second-order language. In general, an $(n+1)$th-order language may be used as a metalanguage for a semantics for an nth-order language.[3]

This section begins with a brief account of Williamson's most important

[2]This view is also defended in Rayo and Williamson (2003).

[3]As a matter of fact, we are only forced to the next order, $n+1$ say, if the object language contains predicates taking variables of order n.

arguments for using higher-order metalanguages. One is derived from WA and one from the notion of strict adequacy. In Section 5.1.2, we give a variant of Williamson's definition of truth for first-order object languages. The definition involves second-order quantification and thus raises questions about the range of the second-order variables. In Section 5.1.3 we briefly consider Quine's critical view on second-order logic, i.e., that the second-order quantifiers really are first-order, ranging over sets. That view threatens to collapse Williamson's hierarchical approach and we explore his defence together with some of its consequences in Section 5.1.4.

5.1.1 WA and strict adequacy as arguments for higher-order metalanguages

The argument from WA for the appropriateness of a second-order semantics is an indirect argument, showing that first-order languages are ill-suited for semantic theorising in ways that second-order languages are not. For, given that we use our first-order quantifiers unrestrictedly, WA shows, according to this argument, that first-order quantification over interpretations is paradoxical. Since we standardly quantify over interpretations in semantic theories, it follows that WA rules out first-order languages as vehicles of semantic theorising. In particular it rules out a first-order Tarskian definition of logical truth and consequence in terms of interpretations. Indeed, if we merge interpretations and assignments as in Section 4.1 it also rules out a Tarskian definition of truth.

Although the argument from WA is indirect, Williamson's view that we should use a second-order metalanguage for unrestricted first-order quantification is not merely a normative standpoint, telling us what is the best practice when defining truth for absolute first-order quantification. Instead, once interpretations are properly understood they are to be completely shunned as first-order quantifiable entities; when generalising over them, Williamson explains, we quantify into predicate position rather than name position.

Accordingly, Williamson suggests that we should define truth in relation to *denotational relations* rather than first-order quantifiable interpretations. Denotational relations are the second-order analogues of interpretations.

Thus, according to this idea, the problematic first-order quantification over interpretations is replaced by second-order quantification over relations.

To get a grip on what relations count as denotational, assume that $\mathscr{L}^m$ is a second-order metalanguage of some first-order object language $\mathscr{L}$. For simplicity, assume that $\mathscr{L}$ lacks function symbols and that all its predicates are monadic. A dyadic relation D is now said to be a *denotational relation for* $\mathscr{L}$ if it relates each singular term t (including the first-order variables) of $\mathscr{L}$ to one, and only one, object, and relates each predicate P of $\mathscr{L}$ to zero, or more, objects.[4] The idea is then to define, in $\mathscr{L}^m$, a truth predicate 'Tr' relative to denotational relations for $\mathscr{L}$ by recursion on the complexity of formulas. A predicate is defined as true of an object according to a denotational relation D if D holds between the predicate and the object in question.

Note that, while 'Tr' is an $\mathscr{L}^m$-predicate, 'P' and 't' are variables in $\mathscr{L}^m$ ranging over $\mathscr{L}$-predicates and $\mathscr{L}$-terms respectively. We will also use 'v' as a meta-variable ranging over the variables of $\mathscr{L}$.

One important feature of this type of semantics that Williamson stresses is that an $\mathscr{L}$-predicate need not be taken to denote a *set* of objects. Nor do we have to assume that a semantic value of a predicate is any other kind of object that somehow collects many things into one. For, rather than denoting the collection of objects of which it is true, a predicate is interpreted by a denotational relation D to denote each object of which it is true according to D. One might think that this is a mere triviality since the image of a relation R from an object o, in this case a predicate, is normally regained by means of the set $\{\, x \mid R(o, x)\,\}$. However, we are not always at liberty to assume that this is possible. We may want to interpret a predicate as being true of a collection of objects that, according to standard set theory, do not form a set, e.g., the class of all sets. And indeed, from a second-order perspective there is *prima facie* nothing that prevents the adoption of a denotational relation making a predicate true of all sets.

This actually illustrates the second of the two main arguments that Williamson presents for turning to second-order semantics. It starts from the quite plausible idea that our semantics should not restrict our ways of in-

[4] Polyadic predicates would stand in the relation D to sequences of objects.

terpreting predicates. That is to say, given a predicate *P* and a semantics S, for any possible interpretation of *P*, we want there to be an interpretation interpreting *P* accordingly in S. A semantics meeting this criterion is said to be *strictly adequate*.[5]

To formulate the criterion of strict adequacy in a clear and distinct way is quite laborious. A first attempt may be the formulation of the first premise of WA:

> [...] whatever contentful predicate we substitute for 'F', some legitimate interpretation (say, I(F)) interprets the predicate letter P accordingly:
>
> For everything o, I(F) is an interpretation under which P applies to o if and only if o Fs. (Williamson, 2003, p. 426)

Thus formulated, in terms of substitution of predicates, it becomes highly dependent on what contentful predicates are available in our metalanguage. Since there may be more interpretations of predicates than there are contentful predicates, this formulation is ill suited as a *sufficient* criteria for strict adequacy. But since each contentful predicate of the metalanguage reasonably corresponds to a possible interpretation of predicates in the object language, this version of strict adequacy constitutes a *necessary* criterion.

Here is an attempt at a general formulation of a sufficient criterion:

> Say that a semantics based on ϒ-interpretations is *strictly adequate* for a language *L* only if every semantic value that a non-logical expression in *L* might take is captured by some ϒ-interpretation. (Rayo and Williamson, 2003, p. 353, italics in the original)

Here, talk of contentful predicates is replaced by talk of semantic values of the non-logical expressions of the object language. Presumably there are no contentful predicates lacking semantic values. If so, the semantic version entails the syntactic version.

Both versions may be used to argue for the need of higher-order metalanguages. The substitutional version is used in WA to argue for higher-order

[5] See, e.g., Rayo and Williamson (2003) or Rayo (2006).

semantics. The semantic version opens up for what we may call the cardinality argument for higher-order languages. The idea is to use some version of Cantor's theorem, i.e., the cardinality of the power set of a set is strictly greater than the cardinality of the set. In the set-theoretic version we may argue as follows. Assume that interpretations are individuals and that the semantic values of monadic predicates are sets of individuals. Then the number of semantic values of predicates is strictly greater than the number of interpretations, which has to be less than or equal to the total number of individuals. It follows that interpretations cannot be constructed as first-order objects in a strictly adequate semantics.[6]

Rayo and Williamson (2003) introduce the concept of strict adequacy when arguing that the semantics based on interpretations formulated in a metalanguage containing *second-level predicates*, i.e., predicates that take second-order variables in their argument places, is not sufficient for providing a strictly adequate semantics for object languages containing second-level predicates. More generally, say that a language with *n*th-order quantification, but with no predicates taking *n*th-order variables in their argument places, is a *basic nth-order language*. If, in addition, an *n*th-order language has predicates with *n*th-order variables in their argument places, call it a *full nth-order language*. It is then possible to show that a basic (full) *n*th-order language cannot be given a strictly adequate semantics in a basic (full) *n*th-order language. But a basic *n*th-order language may be given a strictly adequate semantics in a full *n*th-order language, and a full *n*th-order language may be given a strictly adequate semantics a basic $(n + 1)$th-order language.[7]

First-order languages may be given a strictly adequate semantics within a basic second-order language according to the above, but not within a first-order language. This is thus the third way the requirement that a semantics for absolute quantification ought to be strictly adequate provides an argument for turning to second- or higher-order languages. Just how strong these arguments are hinges on how prone one is to accept the requirement that a semantics ought to be strictly adequate. We return to that question

[6] See Rayo (2002, 2006) for formulations of this argument when predicates are interpreted as pluralities.

[7] See Rayo and Uzquiano (1999) and Rayo (2006).

in Section 5.3.

To summarise, Williamson's suggestion of using a second-order metalanguage to construct a semantics for first-order languages is motivated, first, by the immunity of such a semantics to WA and, secondly, by its ability to provide a strictly adequate semantics.

5.1.2 Williamson's truth definition

In this section we turn to Williamson's truth definition, or rather, a variant of his definition. It takes the form of an inductive definition of the $\mathscr{L}^m$-predicate '*Tr*'. We also briefly consider why WA cannot be derived from this definition and take notice of a pitfall that, Williamson warns, might lead back to paradox if we diverge from the strictly higher-order perspective.

The variant of Williamson's definition presented here draws heavily on the original presentation given in Williamson (2003, Section X), both in notation and content.

Let us establish some convenient notational abbreviations. If D is a denotational relation for $\mathscr{L}$, let D‘t be the x such that $D(t, x)$, for any term t in $\mathscr{L}$.[8] Furthermore, to deal with quantified $\mathscr{L}$-sentences, let:

$$D_{[v/x]}(y, z) \leftrightarrow (y \neq v \wedge D(y, z)) \vee (y = v \wedge z = x)$$

Thus, $D_{[v/x]}$ is just like D with the (possible) exception that it relates the variable of $\mathscr{L}$ which v names to x rather than D‘v.

We may now give the following version of Williamson's definition of a formula φ being true according to D:[9]

Definition 5.1.1. Let φ be a formula of $\mathscr{L}$ and D a denotational relation of $\mathscr{L}^m$. Then φ *is true according to D*, $Tr(D, \varphi)$, if and only if:

1. φ is $P(t)$ and $D(P, D\text{‘}t)$; or
2. φ is $t_i = t_j$ and D‘t_i is identical to D‘t_j; or
3. φ is $\neg\psi$ and not $Tr(D, \psi)$; or

[8] Recall our use of 'P' and 't' as variables in $\mathscr{L}^m$, ranging over $\mathscr{L}$-predicates and $\mathscr{L}$-terms respectively, and 'v' as ranging over the variables of $\mathscr{L}$.

[9] For simplicity, $\mathscr{L}$ is assumed to contain no function symbols.

4. φ is $\psi \wedge \chi$ and both $Tr(D, \psi)$ and $Tr(D, \chi)$; or

5. φ is $\forall v \psi$ and for all x, $Tr(D_{[v/x]}, \psi)$[10]

If quantification in $\mathscr{L}$ is over absolutely everything, then 5.1.1 is not adequate unless quantification in $\mathscr{L}^m$ is equally over absolutely everything. On the other hand, given such an understanding of $\mathscr{L}^m$, if quantification in $\mathscr{L}$ is over less than everything, then this definition would be inadequate unless we somehow make it possible to restrict the range of quantification in the fifth clause. One way to do that is to specify a domain of quantification. In this kind of semantics, this need not amount to the introduction of sets or set-like entities in the semantic theory. Rather than restricting the fifth clause by some set we may impose a criterion for belonging to the range of quantification. A generic way of doing this is to expand the concept of denotational relation in such a way that D must have, in addition to predicates and singular terms, the universal quantifier in its domain. Then we may treat all objects x such that $D(\text{`}\forall\text{'}, x)$ as falling within the range of the $\mathscr{L}$-quantifiers according to D, and impose this condition as a criterion in the clause mentioned.[11]

The predicates '*Tr*' and '*denotational relation*'—abbreviated '*DR*'—are both second-level predicates in $\mathscr{L}^m$ by taking second-order terms as arguments. This is of some importance when considering the derivability of WA in this type of semantics.

The upshot of WA in the present setting is the substitutional version of strict adequacy, i.e., that the following principle holds, for any monadic $\mathscr{L}$-predicate P, whatever monadic predicate of $\mathscr{L}^m$ we substitute for 'F':[12]

$$\exists D(DR(D) \wedge \forall x(Tr(D_{[v/x]}, P(v)) \leftrightarrow F(x)))$$

By means of the first clause of Definition 5.1.1 we may simplify this into

[10] This definition deviates only slightly from Williamson's. While we define a dyadic predicate '*Tr*' taking dyadic relations and formulas as arguments Williamson defines a monadic predicate 'Tr_D', or 'true-D' as he prefers to call it, treating D more like an index. Nothing of significance will be derived from this difference.

[11] See Rayo and Uzquiano (1999) or Rayo and Williamson (2003).

[12] To avoid mistakes of use and mention one usually use Quine corners, or some similar device, in contexts like these. To enhance readability, Quine corners are not used here. Whether a symbol is used or mentioned should be clear from the context.

$$\exists D(DR(D) \land \forall x(D(P,x) \leftrightarrow F(x)))$$

The next step in WA consists in defining a predicate R such that, for all x, R is true of x if x is not an interpretation under which some specific monadic predicate of $\mathscr{L}$, P say, is true of x. The proxy of 'interpretation' in $\mathscr{L}^m$ is 'denotational relation', and if we try to formalise the definition of R head on we end up with something like the following:

$$\forall x(R(x) \leftrightarrow \neg(DR(x) \land Tr(x_{[v/x]}, P(v)))$$

But this is not a well-formed formula of $\mathscr{L}^m$ since x occurs both as a second-order variable in '$DR(x)$' and a first-order variable, e.g., in the second occurrence in'$x_{[v/x]}$'. This has made some, e.g., Linnebo (2003), suggest that WA is blocked because the definition of R is not well-formed.

However, this head on formalisation of R's definition need not be the only possible route to contradiction. Williamson notes that one way to make the definition of R well-formed is to use, what we may call, a type-lifting operator. Although, our setting of the second-order semantics differs from Williamson's, it is still possible to give an argument in his spirit.

Thus, let f be a function from objects to denotational relations, and define

$$\forall x(R(x) \leftrightarrow \neg(DR(f(x)) \land Tr(f(x)_{[v/x]}, P(v)))$$

which is equivalent to

$$\forall x(R(x) \leftrightarrow \neg(DR(f(x)) \land f(x)(P,x)))$$

Substitution gives

$$\exists D(DR(D) \land \forall x(D(P,x) \leftrightarrow \neg(DR(f(x)) \land f(x)(P,x)))$$

Let D_R be a witness of the existential quantification. Then, clearly, if we can find an object r such that $f(r) = D_R$, we would have a contradiction. But this is simply a *reductio ad absurdum* of the existence of such an r, i.e. we have that $\neg\exists x(D_R = f(x))$. Hence, WA is blocked.

Thus it seems that the contradiction in WA cannot be derived in this kind of semantics. But there are other pressing challenges to be addressed.

In standard model-theoretic semantics the assumption of absolute quantification poses well-known problems. Not only is there the problem of representing the domain of quantification, but also, since monadic predicates have sets as their semantic values, perfectly sound definitions such as

$$P(v) \leftrightarrow v = v$$

must fail if v may take any value without restriction. That is to say, P would then have to be interpreted as a proper class.

These problems reappear in second-order semantics if second-order languages are interpreted by means of standard model-theoretic semantics based on standard set theory. For given such an understanding of the second-order metalanguage the domain and semantic values of predicates will still be sets. Thus, accepting 5.1.1 as an adequate definition of truth for $\mathscr{L}$, we have merely elevated the problem of interpreting languages with absolute (first-order) quantification from first-order languages to second-order languages.

Williamson further notes that if we interpret our second-order metalanguage in a first-order meta-metatheory, e.g., set theory, WA may reappear. To avoid this he suggests that we should climb the hierarchy of higher-order languages when providing semantics for the metalanguages, meta-metalanguages, and so on.

> Generality-absolutists may fall back into paradox if they somehow commit themselves to a reduction of quantification into predicate position to quantification into name position, by giving the semantics of the second-order meta-language in a first-order meta-meta-language or otherwise. But if they stick resolutely to the higher-order viewpoint throughout the hierarchy of meta-languages, they can avoid paradox. In particular, they should reject Quine's insistence on regimenting a theory into a first-order language as the test of its ontological commitments. (Williamson, 2003, p. 454–455)

5.1.3 From a Quinean point of view

Before we proceed to the details of Williamson's higher-order account, it is helpful to rehearse some of the ingredients of Quine's ideas of ontological commitment and his critique of higher-order quantification.

Consider the following telling passage:

> To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable. [...] The variables of quantification, 'something', 'nothing', 'everything', range over our whole ontology, whatever it may be; and we are convicted of a particular ontological presupposition if, and only if, the alleged presupposition has to be reckoned among the entities over which our variables range in order to render one of our affirmations true. (Quine, 1948, p. 13)

The variables spoken of in this passage are not explicitly required to be first-order. However, this becomes clear once we appreciate Quine's critique of second-order quantification.

The elaboration on this critique in the often quoted Quine (1970) begins by an explanation of the fallacy of confusing the schematic use of predicates with talk about predicates and what predicates may be taken to denote:

> Consider first some ordinary quantification: '$(\exists x)(x$ walks)', '$(\forall x)(x$ walks)', '$(\exists x)(x$ is a prime)'. The open sentence after the quantifier shows 'x' in a position where a name could stand; a name of a walker, for instance, or of a prime number. The quantifications do not mean that names walk or are prime; what are said to walk or to be prime are things that could be named *by* names in those positions. To put the predicate letter 'F' in a quantifier, then, is to treat predicate positions suddenly as name positions, and hence to treat predicates as names of entities of some sort. The quantifier '$(\exists F)$' or '$(F$)' says not that some or all predicates are thus and so, but that some or all entities of the sort named by predicates are thus and so. The logician who grasps this point , and still quantifies 'F', may say that these entities are attributes; attributes are for

> him the values of '*F*', the things over which '*F*' ranges. The more confused logician, on the other hand, may say that these entities, the values of '*F*', are predicates. He fails to appreciate the difference between schematically *simulating* predicates and quantificationally talking *about* predicates, let alone talking about attributes. (Quine, 1970, p. 66–67, italics in the original.)

Quine continues by rejecting the idea that attributes, or concepts, may be considered possible values of '*F*' mainly because they have unclear identity criteria. Sets, on the other hand, are individuated by the principle of extensionality, and may thus be suggested as an alternative to attributes as values of '*F*'. But thus understood, Quine explains, quantification of '*F*' would just be a deceptive way of writing set theory:

> But I deplore the use of predicate letters as quantified variables, even when the values are sets. Predicates have attributes as their "intension" or meanings (or would have if there were attributes), and they have sets as their extensions; but they are names of neither. Variables eligible for quantification therefore do not belong in predicate positions. They belong in name positions. (Ibid.)

That is to say, instead of writing '*F*' Quine suggests that we should write 'x has y' in case of '*F*' taking attributes as values, and '$x \in y$' if the values are sets. In both cases quantification into the position of '*F*' would be reduced to first-order quantification. But that is precisely the type of reduction that Williamson wants to reject in order to avoid paradox in the meta-meta-language.

5.1.4 The higher-order point of view

Williamson's recipe to avoid the Quinean reduction by systematically using higher and higher-order languages clearly imposes a need to explain over what the higher-order variables are supposed to range. But such an explanation is not immediately at hand. Williamson is sometimes considered a conceptualist on a par with Frege.[13] A conceptualist in that sense understands the second-order variables as ranging over concepts of individuals, third-order variables over concepts of such concepts, and so forth. However, such a formulation of what the variables range over runs into familiar problems. We might, for instance want to deny that concepts are individuals by uttering "concepts are not individuals", which seems to entail, since 'concept' here takes the place of a concept of individuals, that concepts are individuals after all.

> What is really needed is a higher-level predicate that stands to first-level predicates as the first-level predicate 'is an object' stands to names. For the same reason, the attempt to contrast objects and concepts as saturated and unsaturated respectively is deeply misleading, for 'unsaturated' is the negation of 'saturated' and the two adjectives belong to the same grammatical category; but whereas 'is saturated' is a first-level predicate, we need a higher-level predicate in place of 'is unsaturated' to do the required work. The distinction must remain one of grammar and not of ontology, because one cannot use first-level and second-level expressions in the same grammatical context to articulate an ontological distinction without violating constraints of well-formedness. (Williamson, 2003, p. 458.)

Williamson continues to explain what this view implies for the suggested account of quantification:

> On this view of second-order quantification, we must reject as misconceived the questions 'What does quantification into predicate position quantify over?' and 'What are the values of variables in predicate position?'; in particular, we must not answer 'Concepts'. If second-level analogues of those questions make sense, the answer to them will be the second-level analogue of 'Objects' as an answer to the questions 'What does quantification into name position quantify over?' and 'What

[13] See, e.g. (Linnebo, 2006, Section 6.3).

> are the values of variables in name position?' (Williamson, 2003, p. 459.)

This reveals a tension between higher-order semantics and natural language. Recall Tarski's idea that natural languages possesses the property of universality: 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language'.[14] Let us make a somewhat weaker version of this a working hypothesis:

WU If we can speak meaningfully about anything in an interpreted formal language, we can also speak about it in natural language.

The tension now consists in the shortage of second-level predicates in natural languages. We may, of course, claim that a given symbol of $\mathscr{L}^m$ is to be used as the second-level analogue to 'Objects', but there seems to be no natural language expression for paraphrasing such a symbol. Indeed, there seems to be very few second-level predicates in natural language, and even fewer, if any, of higher levels. Thus Williamson explains that:

> Perhaps no rendering in natural language of quantification into predicate position is wholly satisfactory. If so, that does not show that something is wrong with quantification into predicate position, for it may reflect an expressive inadequacy in natural languages. We may have to learn second-order languages by the direct method, not by translating them into a language with which we are already familiar. [...] We must learn to use higher-order languages as our home language. Having done so, we can do the semantics and metalogic of a higher-order formal language in a higher-order formal metalanguage of even greater expressive power. (Williamson, 2003, p. 458)

Thus this view entails the negation of **WU** and its plausibility will be inversely proportional to the plausibility of that hypothesis. For those prone to accept **WU**, Williamson's proposal will thus appear to be very bold.

[14] See page 41.

We return to Williamson's conceptual approach in Section 5.3. There we discuss some familiar problems concerning concepts that apparently cannot be restricted to some one type. This kind of problem was originally noted by Russell, e.g., in (1908; 1910), and has been further discussed by Linnebo (2006). We also present a new dilemma for the motivation of using a type-theoretic framework.

Let me close this section with a brief note on the requirement that the distinction between concepts and individuals, as well as the distinctions between concepts of different orders, must belong to grammar rather than ontology. This is not to say that the conceptual approach is merely formalistic. The concepts partitioned into types form ranges of the higher order variables even though this fact may be impossible to express properly in natural language. But such a view on concepts raises the question of their precise nature. This need not be a problem, but giving up natural language as our home language for semantic theorising makes a fully worked out position hard to formulate. Any alternative avoiding this implication is thus preferable. In the next section we take a look at an attempt at using plurals as an alternative to the conceptual approach.

5.2 Rayo's plural logic approach

George Boolos (1984; 1985) suggests that we may use plural logic to interpret monadic second-order logic. The advantage, Boolos argues, is that we get the semantics for monadic second-order logic without introducing any entities in addition to those already assumed by the use of first-order quantifiers. In particular we need not assume classes or concepts among the values of the second-order variables.

In light of the preceding section a natural suggestion would be to employ Boolos's ideas by using plural logic to interpret the second-order metalanguage of Definition 5.1.1. Again the advantage would be the availability of a second-order language at the price of a first-order ontology. However, we cannot use Boolos's account without alteration. For one thing, the language of Definition 5.1.1 is not a *monadic* second-order language. Thus, to get the idea off the ground, a more general version of Boolos's account is needed.

In Section 5.2.1 we recall some details of Boolos's account of plurals and monadic second-order languages. We proceed in Section 5.2.2 to consider a few critical points from Williamson (2003) concerning plural logic as a metatheory. To some extent, these points form a minimal requirement for any theory of plurals to construct a semantics for absolute quantification. In Section 5.2.3 we review Agustín Rayo's (2006) proposal to provide such a metatheory, and a fully stated semantics, as a generalisation of Boolos's plural logic.[15]

5.2.1 Boolos on plurals

Boolos's project of ridding monadic second-order logic from classes or concepts may be considered an application of Russell's version of Occam's razor:

> Entities are not to be multiplied without necessity. (Russell, 1914, p. 112)

In fact, Boolos expresses the matter in the same way, save for one word, though without referring to neither Russell nor Occam:

> Entities are not to be multiplied beyond necessity. (Boolos, 1984, p. 72)

The omission of providing the historical references may be for a good reason. While Russell refers to Occam's razor when replacing entities by logical constructions of other entities in order to reduce the number of primitives needed to explain the external world, Boolos uses the principle in a much more drastic way. Thus, he does not try to replace the values of second-order

[15] It should be said that Rayo, in recent texts, no longer defends absolutism, but argues that the assumption of absolute quantification rests on a misconception of the relation between our language and reality. The core of the argument in Rayo (2012), as I understand it, is that the metaphysical reality may be represented in a variety of ways, and this may give rise to a number of different totalities of quantifiable "objects". An all-inclusive domain would have to include all objects of each of these possible totalities. But, since the notion of representation is not sufficiently constrained to provide a definite collection of the totalities of quantifiable objects, the conception of an all-inclusive domain is defective. See also Rayo (2017).

variables with constructions from values of first-order variables—such an approach would presumably lead back to classes, concepts or something similar—but argues that *no* additional entities, constructed or not, besides the values of first-order variables are required by either monadic second-order statements, or statements containing plural quantification. Here is how he sums up the findings of his (1984):

> The lesson to be drawn from the foregoing reflections on plurals and second-order logic is that neither the use of plurals nor the employment of second-order logic commits us to the existence of extra items beyond those to which we are already committed. We need not construe second-order quantification as ranging over anything other than the objects over which our first-order quantifiers range [...] Ontological commitment is carried by our *first*-order quantifiers; a second-order quantifier needn't be taken to be a kind of first-order quantifier in disguise, having items of a special kind, collections, in its range. (Boolos, 1984, p. 72, italics in the original.)

These conclusions rest mainly on two results, one technical and one philosophical. The technical result is the interpretability of monadic second-order logic in a fragment of English that contains plural quantification and predication.[16]

The philosophical result concerns the ontological commitments induced by plural quantification, i.e., that it requires nothing in addition to whatever is within the range of the first-order quantifiers.

Boolos's argument for this result starts with a critique of Quine's ideas on ontological commitment, and in particular the two ideas that the ontological commitment of a theory may be revealed only if the theory is regimented in a first-order formalism and that second-order variables must range over classes of whatever the first-order variables takes as values.[17]

To illustrate that the requirement of a first-order regimentation is unwanted, Boolos notices that there are examples of English sentences that

[16] The converse is also true, i.e., that the fragment identified by Boolos is interpretable in monadic second-order logic, but this is of lesser importance here.

[17] See quote on page 96 and Boolos (1985, p. 77).

arguably cannot be regimented in a first-order framework. One such sentence is the Geach-Kaplan sentence:

(1) Some critics admire only one another.

If we assume that the universe of discourse consists solely of critics, (1) has a rather straightforward *second*-order regimentation:

(2) $\exists X(\exists x X(x) \wedge \forall x \forall y(X(x) \wedge A(x,y) \rightarrow x \neq y \wedge X(y)))$.

But there is no first-order equivalent to (2).

Boolos credits Kaplan for the method of showing this. Find an instance of (2) that is true in all and only the non-standard models of arithmetic. This may be achieved by putting '$x = 0 \vee x = y + 1$' for A. Then, given any non-standard model, we may take X to be the class of all the non-standard elements in the model. That the instance is false in the standard model is seen if we let X be any set of natural numbers. Then X has a least element x. If x is 0, then put $y = 0$ and otherwise let $x = y + 1$. Since x is least, y cannot belong to X. Now, if there was a first-order sentence φ equivalent to the instance of (2) under discussion, its negation would be a first-order characterisation of the standard model of arithmetic. Hence there is no such φ.[18]

Examples like (1), which requires second-order quantification for its regimentation, make Boolos suggest that we should reject Quine's idea that ontological commitment requires a first-order regimentation.

It is not immediately clear, however, that such a liberalised view on regimentation helps. (1) has a straightforward second-order regimentation in (2), but, according to Quine, second-order logic is just a misleading way of writing set-theory. Thus (3) would be a more perspicuous formulation of (2):

(3) $\exists \alpha(S(\alpha) \wedge \exists x(x \in \alpha) \wedge \forall x \forall y(x \in \alpha \wedge A(x,y) \rightarrow x \neq y \wedge y \in \alpha))$

where 'S' is supposed to hold for all sets. Now, (3) is first-order and, by

[18] See Boolos (1984, p. 57).

Quine's standards, we would be committed to sets when asserting (1).

But, according to Boolos, this is obviously an unwanted consequence. We shouldn't be committed to sets by asserting (1). For instance, it would be absurd if a hard-headed nominalist that shuns abstract objects in general, and perhaps sets in particular, argued for the falsity of (1) from its commitment to sets.[19] Having said that, Boolos needs to reject that regimentation ought to be first-order together with Quine's set-theoretic reading of second-order formulas.

That (1) cannot be regimented in first-order logic together with the observation that it nevertheless seems to involve quantification over first-order entities rather than second-order entities such as sets or properties of individuals, indicates that there are two distinct types of first-order quantification over individuals in English, namely ordinary singular quantification and quantification into plural noun position, or plural quantification for short.

That it is the occurrence of plural quantification which causes the impossibility to give certain sentences a first-order regimentation is nicely illustrated by Boolos in the following pair of sentences:

(4) There is a horse that is faster than Zev and also faster than the sire of any horse that is slower than it.

(5) There are some horses that are faster than Zev and also faster than the sire of any horse that is slower than them.

These sentences differ only by (5) being a plural variant of (4). Yet (4) may be formalised in first-order logic, while (5) cannot.[20]

Thus, rather than providing arguments for second-order regimentation in the search for the ontological commitment of sentences, considerations of (1) and similar sentences are taken to vindicate Boolos's intuitions that the plural quantification they involve ranges only over individuals. This

[19] See Boolos (1985, pp. 76–77).

[20] Again we may prove this by giving a second-order regimentation of the sentence involving plural quantification that is true in all non-standard models of arithmetic but false elsewhere. See Boolos (1984, pp. 57–58) for details.

neatly comes to use when Boolos turns to his main concern, viz. second-order set theory. Without introducing proper classes, how is it possible to claim, he asks, that $\exists X \forall x(Xx \leftrightarrow x \notin x)$, an instance of the valid $\exists X \forall x(Xx \leftrightarrow Ax)$, is true when our first-order quantifiers range over all sets? His answer is to employ plural quantification:

> Abandon, if one ever had it, the idea that use of plural forms must always be understood to commit one to the existence of sets (or "classes," "collections," or "totalities") of those things to which the corresponding singular forms apply. The idea is untenable in general in any event. There are some sets of which every set that is not member of itself is one, but there is no set of which every set that is not a member of itself is a member, as the reader, understanding English and knowing some set theory, is doubtless prepared to agree. Then, using the plural forms that are available in one's mothers tongue, translate the formulas into that tongue and see that the resulting English (or whatever) sentences express true statements. (Boolos, 1984, p. 66)

The suggested translation * for second-order set theory is as follows.

- $(Xx)^* =$ "it$_x$ is one of them$_X$"
- $(x \in y)^* =$ "it$_x$ is a member of it$_y$"
- $(x = y)^* =$ "it$_x$ is identical to it$_y$"
- Translation commutes with the propositional connectives as usual.

Furthermore, we restrict attention to formulas Φ that have no vacuous quantifiers and no quantifier within the scope of a quantifier with which it shares its variable. Let $\Phi_{[\varphi/\psi]}$ be the result of replacing each occurrence of φ in Φ by ψ. Then,

- $(\exists x\Phi)^* =$ "there is a set that$_x$ is such that Φ^*"
- $(\exists X\Phi)^* =$ "either there are some sets that$_X$ are such that Φ^* or $\Phi_{[Xx/x \neq x]}$"

The translation is slightly complicated by the existential import the plural reading of '$\exists X$' invokes. If we read '$\exists X$' as 'there is a class X such that' this does not in general entail that the class should be non-empty. On the other hand, the plural reading, 'there are some things$_X$ such that,' seems to entail the existence of at least one such thing. Hence Boolos suggests that a formula $\exists X \Phi$ should be translated with the extra disjunct added where occurrences of Xx are replaced by '$x \neq x$'.[21]

If this semi-formal fragment of English really is ontologically innocent, as Boolos claims, then arguably, monadic second-order logic may be taken as carrying no ontological commitment in addition to whatever its first-order variables take as values. To argue for that conclusion Boolos relies on our intuitions about natural language sentences involving plural quantification. These intuitions may be right, of course, but other authors, notably Resnik (1988) and Parsons (1990), have expressed contrary intuitions. The existence of conflicting intuitions regarding plural quantification makes it hard to reach a decisive conclusion about its ontological commitment.[22] It is rather clear that, such intuitions about natural language are not sufficient to show that the version of plural logic used in this chapter brings about no new ontology. Thus, we shall not enter this debate here, but tentatively accept Boolos views in the remainder of this chapter.

It is important to be clear about the structure of Boolos's argument. There are two types of quantification over individuals in English, singular and plural quantification. We may interpret second-order logic in plural logic (and vice versa). Hence we need not assume the existence of sets or classes to interpret monadic second-order statements. Simply interpret such statements by means of the plural devices given in English. Of course, we may extrapolate any such argument to any pair of logics, given that, at least, one is interpretable in terms of the other. Such an approach will be interesting only if there are external arguments showing that the languages ultimately commit us to different ontologies. In Boolos's case such external arguments come from inspection of certain natural language sentences and our intuitions concerning them. Even if one is hesitant regarding the valid-

[21] See Boolos (1984, p. 68).
[22] See Linnebo (2003, 2012) for rewarding discussions about this theme in the philosophy of plural logic.

ity of Boolos's findings regarding these sentences, it is quite possible to accept the ontological innocence as a working hypothesis.

A natural idea, we said, is now to put the plural reading, or some more general alternative of it, to work when quantifying into the position of D in Definition 5.1.1. For, if Boolos is right, the adoption of a second-order metalanguage need not impose additional ontology. Furthermore, WA can be avoided.

5.2.2 Williamson's critique of the plural approach

Indeed, Williamson does consider the possibility of using some general version of Boolos's plural reading to interpret his suggested metalanguage but concludes, having identified five problematic points, that it is in fact not satisfactory.[23] Thus Williamson points out that second-order monadic variables stand in predicate position, whereas they get translated into name position.[24] This indicates that plural quantification differs in kind from second-order quantification. He also argues that the need for the extra clause for handling predicates that apply to nothing "gives [a] hint of Procrustean activity" (Ibid. p. 456). There is also a problem with quantification into polyadic predicate places since there seems to be no natural language reading corresponding to such quantification.[25] The fourth point is that a plural reading of second-order quantification in modal contexts would, since the plural variable is naturally rigid, make the second-order variables rigid as well.

The most important of Williamson's points is the fifth:

> A fifth point is that the plural reading has no natural generalization to nth-order quantification for n greater than two. But, by a generalization of the Russellian paradox, we need to use $(n+1)$th-order quantification in the meta-language to define logical consequence for an nth-order object-language that contains nth-level non-logical predicates. For example,

[23] See Williamson (2003, pp. 456–458).

[24] For a discussion, see Simons (1997), which is also the reference Williamson gives.

[25] However, see Rayo and Yablo (2001) for a discussion.

> the formula Π(F) in which the second-level predicate Π is applied to the first-level predicate 'F' might have the reading 'the things that F collectively lifted a piano'. A Tarskian definition of logical consequence for a second-order language with the second-level predicate P in the spirit of the account above would involve third-order quantification into the position of P, for which the plural reading would not suffice. (Williamson, 2003, p. 457)

Note that Williamson merely points out that the plural *reading* does not generalise to higher-order quantification. That is, if first-order existential quantification corresponds to natural language quantification by means of 'something', and second-order quantification corresponds to 'some things', there is no natural language expression corresponding to third-order quantification. In that sense, the plural reading is not available for higher-order languages. But, expressibility in natural language need not be taken as a necessary condition for the intelligibility of higher-order plural quantification.

Moreover, Williamson uses WA to motivate the need of $(n+1)$th-order quantification to define logical consequence for nth-order languages. But in Chapter 4 we argued that WA is inconclusive as an argument against first-order quantification over absolutely everything. Thus, in order for Williamson's fifth point to have force, the hierarchy of higher-order metalanguage needs to be motivated by some other means. Rayo uses the semantic version of strict adequacy to argue for the necessity of the higher-order plural account by means of the cardinality argument. Moreover, while the cardinality argument is taken to show that first-order languages are insufficient for our purposes, Rayo argues that we can provide an adequate semantics for plural languages if we help ourselves to higher and higher levels of plural predication and quantification.

5.2.3 Rayo's account of higher-order plural logic and absolute quantification

Williamson's fifth point concerns the possibility of interpreting the second-order metalanguage in terms of plural quantification and predication. But, if we want a semantics for first-order quantification, the detour over second-order logic may be unnecessarily indirect. That is, rather than using a second-order metalanguage to interpret the first-order languages and then translate the metalanguage into a plural language one may use the plural language as a metalanguage in the first place.

Boolos, as we saw, relies on natural language expressions involving plural locutions to interpret monadic second-order languages. But, as Williamson hints at, we may need nth-order metalanguages for arbitrarily large values of n, and while first- and second-order existential quantification corresponds quite naturally to *something* and *some things*, there are no corresponding natural language expressions for third-, or higher-order quantification. Hence, once we proceed beyond (monadic) second-order languages the available plural quantifiers in natural language no longer give us guidance for how to read the higher-order quantifiers. There is no part of English, more inclusive than the part identified by Boolos, that can be used to read higher-order plural equivalents.[26]

Lacking the natural language reading, an alternative way to get a plural metalanguage is to set forth a *formal* language with higher-order equivalents of plural quantification and argue that it, in fact, can be used for semantic theorising even without natural language correspondents to some of its expressions. Rayo (2006) advocates this possibility suggesting that the semantic categories of a categorial grammar for a given language may be legitimate even if there is no translation into natural language. It might even be the case that such a translation is in principle impossible and that the suggested language is irreducible to natural language. Instead, Rayo suggests,

> [...] a semantic category $\mathcal{C}$ is *legitimate* just in case it is in principle possible to make sense of a language whose semantic properties are accurately described by a categorial semantics employing $\mathcal{C}$. (Rayo, 2006, p. 222)

[26] However, see Linnebo and Nicolas (2008) for a discussion that there are superplural predication in English. This might provide some support that there may be third-order plural quantification. The argument given does not seem to generalise to arbitrarily high orders of plural predication and quantification however.

Of course, this needs further elaboration to become sufficiently clear and distinct. In particular, it needs to be clear precisely what it means to make sense of a language, what semantic properties are required to have an accurate description, what such a description may look like, and how we may decide whether or not it is accurate. But for now, this informal explanation suffices.

The notion of reference is important. Following Rayo we call a predicate 'P' that only takes singular terms as arguments a *first-level predicate*. Rayo argues that, under the assumption that quantification is over absolutely everything, one should reject the idea that the semantic value of 'P' is the set of individuals it is true of. For taking sets as semantic values of predicates implies that predicates such as '…is self-identical' lack semantic value.[27] Rayo continues:

> Rather than taking '…is an elephant' to stand for the set of elephants, I would like to suggest that one should take it to stand for the elephants themselves. It is grammatically infelicitous to say that the semantic value of '…is an elephant' is the elephants. So I shall state the view by saying that '…is an elephant' *refers* to the elephants. (Rayo, 2006, p. 225, italics in the original)

Formally, this is expressed as follows.

(6) $\exists xx(\forall y(y \prec xx \leftrightarrow \textsc{Elephant}(y)) \wedge \textsc{Ref}(\text{'…is an elephant'}, xx))$

(6) extends the fragment of English that Boolos identified by containing a predicate 'REF' that takes a plural term in its second argument place. Thus, besides first-level predicates, i.e., predicates taking only singular terms in their argument places, we need to accept predicates taking plural terms in at least one argument place, i.e., second-level predicates. We add a sequence $i_0, i_2, \ldots, i_{n-1}$ as a superscript to predicates to indicate the level of its arguments. Thus, for instance, $i_j = 1$ if it takes singular individual terms (of level 0) as arguments at place j, and $i_j = 2$ if it takes plural terms (of level 1) as arguments at that place. In case of (6) we get

[27] Rayo (2006, p. 224). We tentatively argued in a similar way on page 94.

$$(7)\ \exists xx(\forall y(y \prec^{1,2} xx \leftrightarrow \textsc{Elephant}^{1}(y)) \wedge \textsc{Ref}^{1,2}(\text{'...is an elephant'}, xx))$$

Monadic second-level predicates, i.e. predicates taking first-level plural terms as arguments, are understood to refer in an analogous way as first-level predicates refer to pluralities. Say that '...are scattered' is a monadic second-level predicate. Rayo claims that we should resist the temptation of employing sets of pluralities as its semantic values:

> I propose instead that the reference of '...are scattered on the table' should be characterised as follows:
>
> $$(8)\ \exists xxx(\forall yy(yy \prec^{2,3} xxx \leftrightarrow \textsc{Scattered}^{2}(yy)) \wedge \textsc{Ref}^{1,3}(\text{'...are scattered'}, xxx))$$
>
> where treble variables are used for *super-plural* terms and quantifiers. There are, of course, no super-plural terms of quantifiers in English, but I would like to suggest that the relevant semantic category is nonetheless legitimate: super-plural quantifiers are to third-order quantifiers what plural quantifiers are to second-order quantifiers.[28] (Rayo, 2006, p. 227, italics in the original)

Note the distinction between 'order' and 'level' in this section. A second-order predicate takes first-order *predicates* as arguments, while a second-level predicate takes first-level *terms* as arguments.[29]

The first-level terms come in two variants: the plural variables, e.g. *xx*, and the result of applying the definite article to a plural noun, e.g., 'the elephants'. Rayo's semantic analysis of first-level terms is on a par with the semantic analysis of first-level predicates, putting

$$(9)\ \exists xx(\forall y(y \prec^{1,2} xx \leftrightarrow \textsc{Elephant}^{1}(y)) \wedge \textsc{Ref}^{1,2}(\text{'the elephants'}, xx))$$

[28] The numbering of the formula has been changed.

[29] This generalises in a straightforward way: An $(n+1)$th-order predicate is a predicate which takes nth-order *predicates* as arguments, while a $(n+1)$th-level predicate is a predicate that takes nth-level terms as arguments.

Thus first-level predicates and first-level terms, though grammatically distinct, share semantic values.

Second-level terms are constructed from second-level predicates in a similar fashion. There is no definite article applying to second- or higher-level predicates, so Rayo introduces a *saturation operator*, 'σ' applying to predicates to form terms. Thus we may form a second-level term $\sigma[P^2(\ldots)]$ from a second-level predicate P^2 such that they are grammatically distinct, but have the same semantic value (which will be a super-plurality).

Proceeding in the same manner Rayo continues to introduce *n*th-level predicates and *n*th-level terms, for each natural number *n*, to get what he calls limit$_\omega$ languages. A limit$_\omega$ language contains

1. Logical connectives: $\wedge, \neg$
2. Variables: v_i^n, for each level $n \geq 0$
3. Individual constants: c_i^0
4. Predicates: P_i^s where s is a sequence of natural numbers indicating how many arguments the predicate in question takes and to what level they belong
5. Logical predicate letters: $=^{1,1}$, $\prec^{n-1,n}$ and Ex^n. 'Ex' is short for exists
6. The term forming operator σ_i^n

The terms of a limit$_\omega$ language are the variables, the individual constants and the result of applying the term forming operator σ^n to formulas φ to get terms $\sigma^n\varphi$ of level $n+1$. Well-formed formulas are defined as one expects: Predicate letters, logical as well as non-logical, apply to terms of appropriate levels in relation to the sequence they have as superscript. Thus for instance $\mathrm{Ex}^3(\sigma^1(\varphi))$ is well-formed (if φ is) saying, roughly, that the super-plurality *xxx* of φ's exist.

Just as in Frege's original notation for quantificational logic, limit$_\omega$ languages treat quantifiers as (higher-level) predicates. But, as Rayo points out, quantification over individuals, pluralities, super-pluralities, etc., may be defined:

$$\exists v_i^n(\varphi) \leftrightarrow \mathrm{Ex}^{n+2}(\sigma^n(\varphi))$$

$$\forall v_i^n(\varphi) \leftrightarrow \neg\exists v_i^n(\neg\varphi)$$

Next, Rayo stratifies the limit$_\omega$ languages:[30]

1. Basic first-level languages: The fragment containing the variables v_i^0 and no others, the logical predicates $=$ and Ex^2, the instance σ_i^0 of the term forming operator but not $\prec$.

2. Full first-level languages: basic first-level languages with non-logical first-level predicates added.

3. Basic $(n+1)$-level languages result from full nth-level languages by enriching them with variables v_i^n, logical predicates $\prec^{n,n+1}$ and Ex^{n+2}, together with the saturation symbol σ_i^n.

4. A full $(n+1)$-level language is the result of enriching a basic $(n+1)$-level language with non-logical predicates of level $n+1$.

In order to make the limit$_\omega$ languages acceptable for semantic theorising about absolute quantification Rayo, by his own standards, has to show, first, that the semantic categories of limit$_\omega$ languages are legitimate, secondly, that the ontology of higher-order plural quantification is the same as the ontology of first-order quantification, and thirdly, that it is possible to provide a strictly adequate semantics for the limit$_\omega$ languages.

Rayo presents an argument that, for each n, the nth-level predicates and terms belong to legitimate semantic categories. The argument assumes that we can quantify over absolutely everything and proceeds by showing that, under that assumption, we cannot translate a second-level language into a first-level language in a truth-preserving way without facing certain difficulties. Rayo admits that the argument only provides preliminary evidence for its conclusion and we shall not enter into the details of it here.[31]

Though some less preliminary argument for the legitimacy of the semantic categories Rayo suggests is surely needed, there seems to be no prima

[30] See Rayo (2006, p. 237). We sometimes ignore the superscripts.
[31] The argument is given in Section 9.5.2 of Rayo (2006).

facie reason for rejecting the possibility that such an argument can be given and that the higher-level terms and predicates may, in the end, be found to be members of legitimate semantic categories. So, let us assume that such an argument has been given. Then, Rayo shows, we may construct a strictly adequate model theory with respect to the categories in question for each level of a limit$_\omega$ language.[32]

Rayo uses the following instance of the semantic version of strict adequacy:

> [A] model-theory for a language L is *strictly adequate* just in case it agrees with one's categorial semantics for L in the following sense: any reference a (non-logical) predicate might take by the lights of one's categorial semantics corresponds to the semantic value the predicate gets assigned by some model of one's model-theory. Thus, given a model-theory whereby the reference of a first-level predicate is a plurality, a model-theory for the relevant language can only be strictly adequate if, for any plurality, there is a model on which a given first-level predicate is assigned a semantic value corresponding to that plurality. (Rayo, 2006, p. 243, italics in the original)

Let L be a first-level language of just one monadic predicate P. A model of that language is then basically a plurality *mm* of pairs of the two forms $\langle P, x\rangle$, where x belongs to the extension of P under *mm*, and $\langle \forall, y\rangle$, where y belongs to the range of the variables in *mm*.

Now, if $M(xx)$ expresses that the plurality *xx* forms a model of L in the sense just outlined, then this semantics is strictly adequate for L if:[33]

(10) $\forall xx \exists mm \forall y((M(mm) \wedge \langle P, y\rangle \prec mm) \leftrightarrow y \prec xx)$

[32] For the technical details of the construction, see the appendix of Rayo (2006).

[33] The types are left out in this formula for readability, but clearly we must treat $\langle P, y\rangle$ as a first-level term. In type-theoretic set theory the Wiener-Kuratovski ordered pair, $\{\{x\}, \{x, y\}\}$, is of type two levels above the type of its constituents x and y. Such an option, treating the ordered pair as a third-level term would make the formula, as well as the notion of model explained, ill-formed. Thus, although Rayo does not explicitly say so, he seems to adopt '$\langle\,,\,\rangle^{n,n}$' as a term forming operator taking terms of level n to ordered pairs of level n.

This version of semantic adequacy may then be seen as an elaboration of the semantic version which was formulated without any reference to particular semantic categories at page 89.

Now, Rayo (2006, p. 244) claims that the *n*th-level languages, for all *n*, basic and full, are ordered with respect to the possibility of providing strictly adequate semantics:

(a) It is impossible to give a strictly adequate model theory for a full *n*th-level language in an *n*th-level language

(b) It is possible to give a strictly adequate model theory for full *n*th-level languages in a basic $(n + 1)$th-level language

(c) It is impossible to give a strictly adequate model theory for a basic $(n + 1)$th-level language in a basic $(n + 1)$th-level language

(d) It is possible to give a strictly adequate model theory for basic $(n + 1)$th-level languages in a full $(n + 1)$th-level language

Thus any language of a finite level can be given a strictly adequate model theory in some language belonging to the hierarchy.

However, as Rayo points out, it follows from (a) and (c), together with the observation that a limit_ω language contains an *n*th-level language as part for each *n*, that it is impossible to give a strictly adequate model theory for a limit_ω language in a limit_ω language. For, assume that φ of some limit_ω language captures the notion of truth-in-a-model. Then φ, having finitely many symbols, is a formula of an *n*th-level language for some *n*, which contradicts the conjunction of (a) and (c).

We are thus faced with two options according to Rayo: either we accept that there are languages for which we, despite their legitimacy, cannot provide an adequate model theory, i.e., we settle for *semantic pessimism*, or we accept an open-ended hierarchy of languages. An open ended hierarchy requires that we give up the limit_ω languages, showing them to be illegitimate somehow, or that we proceed into transfinite levels beyond ω.

Despite known counterarguments Rayo suggests that an open ended hierarchy might be the least unattractive option.

The most important contribution from Rayo's investigations, it seems to me, is not his conclusion that an open-ended hierarchy of languages is preferable, but that there is a real tension between avoiding semantic pessimism and keeping strict adequacy in type-theoretic semantics. That tension may alone serve as an impetus for further investigations of available alternative meta-theories. In the next section we make this impetus even stronger by pointing to some further challenges for both Williamson's and Rayo's approaches.

5.3 Critical remarks on typed semantics

In this section we look at some problems inherent to the typed semantics for absolute quantification. We begin in Section 5.3.1 with the problem of expressing facts, by means of the type-theoretic language, that seem to inevitably cut across types. In Section 5.3.2 we argue that the motivation of typed semantics from semantic adequacy is not as solid as one might expect.

5.3.1 Linnebo's critique

Russell was well aware of the problem of expressing facts about all types within the type-theoretic language itself. His analysis of the paradoxes—the disease for which type theory was meant to be a cure—concludes that they all violate the vicious-circle principle:

> 'Whatever involves *all* of a collection must not be one of the collection'; or, conversely: 'If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total.' (Russell, 1908, p. 225, italics in the original)

But, as Russell notices, adopting the vicious-circle principle has some unwanted consequences:

> The first difficulty that confronts us is as to the fundamental principles of logic known under the quaint name of 'laws of thought'. 'All propositions are either true or false', for example, has become meaningless. If it were significant, it would be a proposition, and would come under its own scope. Nevertheless, some substitute must be found, or all general accounts of deduction become impossible. (Ibid.)

Similar problems apply to the type-theoretic languages suggested by Williamson and Rayo. Linnebo (2006) picks up this theme when he argues that the type-theorists are committed to certain general insights that they will not be able to express. In particular, Linnebo gives the following examples (Linnebo, 2006, pp. 154–155):

- *Infinity.* There are infinitely many different kinds of semantic values.
- *Unique Existence.* Every expression of every syntactic category has a semantic value which is unique, not just within a particular type, but across all types.
- *Compositionality.* The semantic value of a complex expression is determined as a function of the semantic values of the expression's simpler constituents.

The common trait of these examples is that they, in essential ways, require generalisations across types for their intended meanings to get through. But if each quantifier is constrained within some particular type such generalisations are impossible. It follows that the type-theorists cannot express the intended meanings of the general insights that Linnebo claims them to be committed to.

This kind of critique has been met by arguments saying that we may have to live with the lack of a firm point from which we may determinately express such semantic insights that require generalisations across types. Rayo, we said, argues that there are two strategies available. The first strategy is to settle for *semantic pessimism* in the sense that there are, in fact, legitimate languages (e.g. $\mathscr{L}_\omega$) for which we cannot provide a strictly adequate semantics. Furthermore,

> [t]he second strategy is to avoid semantic pessimism by claiming that the *legitimate* languages—the languages it is in principle possible to make sense of—form an open-ended hierarchy such that any language in the hierarchy can be given a strictly adequate model-theory in some other language higher-up in the hierarchy. So there is no legitimate language with respect to which semantic pessimism would threaten. [...] Whatever the details of the hierarchy, what is crucial is that there is no such thing as an *absolute*-level language. (Rayo, 2006, p. 246, italics in the original)

But, as Linnebo's examples show, the postulation of a hierarchy merely leads to yet another form of semantic pessimism, i.e., that certain insights needed for a systematic development of an adequate semantics cannot be expressed. Rayo suggests that we just might have to bite that bullet:

> The postulation of an open-ended hierarchy of languages faces a familiar difficulty: it leads to the result that statements of the form 'the hierarchy is so-and-so' are strictly speaking nonsense. [...]
>
> In spite of its problems, the postulation of an open-ended hierarchy of languages may turn out to be the least unattractive of the options on the table. (Rayo, 2006, p. 247)

Williamson's conceptual hierarchy is challenged by Linnebo's examples in the same way.

Linnebo's examples concern insights that we want to be able to express when developing the semantic theory in the type-theoretic language. The problems of expressing them shows that the development of a semantics for absolute quantification in a type-theoretic language is less straightforward than it first may appear. Thus, facing these insights, the type-theorist finds himself in the position of having, on the one hand, strong arguments for going type-theoretic, and on the other hand, new problems attached to the employment of such languages. If the arguments for turning to type-theoretic semantics seem strong enough a type-theorist may find Rayo's position attractive, i.e., that the type-theoretic languages (in an open-ended hierarchy) is the most plausible option at hand. Thus, if the motives are strong enough, the type-theorist may be willing to accept the inexpressibility of certain insights.

To evaluate the plausibility of a typed semantics in light of the problems that Linnebo and Russell present we need to evaluate the motives for going type-theoretic in the first place. We said that the main arguments for typed semantics are WA, strict adequacy, and, if the semantic version of strict adequacy is employed, the cardinality argument. The attractiveness of the open-ended hierarchy of languages hinges to a great extent on just how convincing one finds WA to be, and how prone one is to accept strict adequacy as a virtue for any semantics for absolute quantification. We discussed WA at length in Chapter 4. In this section we focus on the notion of strict adequacy.

Let us recapitulate the two versions of strict adequacy, i.e., the semantic and the substitutional version, and give them a slightly more precise formulation.

Semantical strict adequacy A semantics is semantically strictly adequate if, for any possible semantic value of a non-logical constant, there is an interpretation interpreting the constant as that value.

Substitutional strict adequacy A semantics is substitutionally strictly adequate if, for any contentful predicate φ, some interpretation interprets a predicate letter P as true of an object whenever φ is true of that object.

These general formulations of strict adequacy may be instantiated in particular semantic frameworks. Thus, for instance, as we pointed out in the previous section,

(10) $\forall xx \exists mm(M(mm) \land \forall y(\langle P^{\langle 1 \rangle}, y\rangle \prec mm \leftrightarrow y \prec xx))$

is an instance of the general formulation of the semantic version of strict adequacy in the framework of plural logic. Or, rather, it is part of such a formulation since it takes into account only semantic values of first-level predicates. There is no way to formulate this principle in its full generality in a type-theoretic framework since the notion of semantic value must always be restricted to a type. Thus, according to the type-theorist's own standards, the notion of semantic strict adequacy is inexpressible.

A similar argument is possible for any instantiation of the substitutional version of strict adequacy. For in its general version it says that for any contentful predicate, *of any type*, there is an interpretation that interprets the predicate accordingly. But such quantification over all contentful predicates violates the type-restrictions.

Thus, both versions of strict adequacy are targeted by the same kind of critique that Linnebo presents regarding the general semantic insights. However, in contrast to those insights, strict adequacy has previously been used to motivate the move to type-theoretic semantics in the first place. Thus, once the type-theorist take the hierarchical language to heart the argument from strict adequacy will no longer be available.

Let us set these matters of inexpressibility aside and consider a problem with the semantic version of strict adequacy. Some instances of that concept seem to lead away from what seems to be the intuitive idea behind it. Consider, for instance, standard model-theoretic semantics as constructed within some set theory. It is semantically strictly adequate since models in standard model theory assign sets to first-order predicates and, for any set, there is some model assigning that set to the predicate in question. In other words, no set is ruled out as a possible semantic value due to the lack of a model assigning it to a given predicate.[34] But, one may argue, standard model-theoretic semantics shouldn't be strictly adequate since it uses words, e.g., *model*, whose intended meaning cannot be the meaning of a predicate expression of the object language according to the suggested semantics. Thus, formulated as above, semantic strict adequacy is too weak to rule out standard model theory as inadequate in the strict sense.

However, that there ought to be some interpretation interpreting a predicate of our object language in accordance with any acceptable predicate of the metalanguage, in this case 'model', is entailed by the substitutional ver-

[34] Thus, given this understanding, the problem with standard model-theoretic semantics is not that it is not semantically strictly adequate. Rather it is the absence of a model representing quantification over absolutely everything that is problematic. The challenge, then, is to find a semantics that is strictly adequate *and* according to which the quantifiers may range over absolutely everything.

sion of strict adequacy. Model-theoretic semantics is then, as formulated within standard set theory, not substitutionally strictly adequate.[35]

In the next section we shall see how the requirement of substitutional strict adequacy adapted to arbitrary well formed expressions of the metalanguage opens up for another critique of the motivation for turning to type-theoretic languages.

5.3.2 A dilemma

That strict adequacy and Linnebo's properties are inexpressible as properties in the kinds of type-theoretic semantics that Williamson and Rayo suggest is an interesting fact in its own right. The inexpressibility of these examples results from the necessity of quantification across types. But there are other kinds of examples as well. In fact, it is possible to derive a contradiction reminiscent of the Grelling-Nelson paradox of heterological words.

Consider the property of *not denoting itself*. Let 'P' abbreviate this property. Then, intuitively, 'P' is true of an object x exactly when x does not denote x. Formally, using the framework of Section 5.1 we put this

$$\forall x(Tr(D_{[v/x]}, P(v)) \leftrightarrow \neg D_{[v/x]}(D_{[v/x]}`v, D_{[v/x]}`v)).$$

But this immediately leads to a contradiction. Instantiation gives

$$Tr(D_{[v/P]}, P(v)) \leftrightarrow \neg D_{[v/P]}(D_{[v/P]}`v, D_{[v/P]}`v).$$

Since

$$D_{[v/P]}`v = P$$

we get

$$Tr(D_{[v/P]}, P(v)) \leftrightarrow \neg D_{[v/P]}(P, P).$$

But, by the truth definition

$$Tr(D_{[v/P]}, P(v)) \leftrightarrow D_{[v/P]}(P, D_{[v/P]}`v).$$

[35] In this case, since not all contentful predicates of the meta-language have semantic values, the semantic version of strict adequacy does not entail the syntactic version.

Using the identity $D_{[v/P]}`v = P$ again we get

$$\neg D_{[v/P]}(P, P) \leftrightarrow D_{[v/P]}(P, P)$$

which is a contradiction.

One might try a number of ways to avoid the contradiction. One could, for instance, argue that there is, for each collection of objects such that $\neg D(x, x)$, another denotational relation, D^*, such that

$$\forall x(Tr(D^*_{[v/x]}, P(v)) \leftrightarrow \neg D_{[v/x]}(D_{[v/x]}`v, D_{[v/x]}`v))$$

But then, 'P' would not really have the intended meaning.

Another way to block the contradiction would be to argue that there cannot be a denotational relation of the sort required. But, since it seems to be unproblematic to assume the existence of a denotational relation D such that a predicate 'Q' is true of all objects x such that $D(x, x)$, and since P is definable in terms of Q in an obvious way, denying the existence of such a D is not immediate. That is, to argue for the non-existence of a denotational relation that interprets a predicate as *does not denote itself*, would also be an argument that, for a given predicate letter 'Q', there cannot be a denotational relation D such that

$$\forall x(D(Q, x) \leftrightarrow D(x, x))$$

Indeed, we could use the above argument just to that end.

When faced with a parallel situation in WA, having defined 'R', Williamson explained that "the naive theorist is committed to treating 'R' as a contentful predicate, since it is well-formed out of materials entirely drawn from the naive theory itself" (Williamson, 2003, p. 126). Thus, if we are justified in rejecting 'D' as a contentful denotational relation, despite it being well-formed out of materials drawn from the type theory itself, one wonders why we are not so entitled in case of 'R'.

This actually constitutes a dilemma, not so much for type-theoretic semantics, as for the way such a semantics is motivated. For if we reject 'D' as a contentful denotational relation, there seems to be nothing that keeps us from ruling out 'R' as a contentful predicate for similar reasons. However, if 'R' is so ruled out, one of the main reasons for turning to type-theoretical semantics in the first place, i.e. WA, becomes neutralised. On the other hand, if we must accept 'R' as a contentful predicate, and thereby restore the strength of WA as an argument for going type-theoretic, then, since no interpretation interprets 'P' as *not denoting itself* on pain of contradiction, the resulting semantics is contradictory if strictly adequate.

The observations in this section are by no means conclusive arguments against the type-theoretic approach, but they motivate further investigations into alternative semantic theories. In the next section we discuss one such suggestion that gives up the type-theoretical approach.

5.4 Linnebo on sets and properties

If we give up the type-theoretic approach to a semantics for absolute quantification there seems to be no other option but to let semantic values of the expressions in the object language be possible values of the first-order variables in the metalanguage. That is to say, whatever we take the semantic values of, e.g., predicate expressions of an object language $\mathscr{L}$ to be, the metalanguage of $\mathscr{L}$ must count the values as some kind of objects. Similarly, interpretations of $\mathscr{L}$ need to be reckoned as objects. Linnebo (2006) shows how this may be accomplished by augmenting standard set theory with urelements, ZFCU, with a theory of properties. Linnebo's idea is then to let properties, rather than sets, be the semantic values of the predicates of $\mathscr{L}$. In this section we take a look at the principal features of this intriguing proposal.

The route to properties goes via the notion of a concept. A concept, Linnebo explains, is "most fundamentally specified by means of some completely general condition" (Linnebo, 2006, p. 159) and two such conditions, $\phi(u)$ and $\psi(u)$ say, determine the same concept if they stand in some suitable equivalence relation Eqv_u. Using second-order lambda calculus, Linnebo (Ibid.) then lays down the following identity criterion for concepts specified by ϕ and ψ:

$$(\Lambda)\quad \Lambda u.\phi(u) = \Lambda u.\psi(u) \leftrightarrow Eqv_u(\phi(u), \psi(u))$$

Thus, when individuating concepts, Linnebo focuses on the conditions rather than on what objects satisfies these conditions. A further characteristic of concepts is that their application conditions remain completely general under the basic algebraic operations of negation, conjunction and existential generalisation. This, Linnebo claims, makes concepts better suited than sets (in ZFCU) as semantic values of predicates.

But concepts, as explained above, are denoted by second-order terms and thus, in order to enable first-order quantification over them they need to somehow be treated as objects. Linnebo suggests that we should nominalise second-order Λ-terms to first-order λ-terms, which are understood as referring to *properties*. He proceeds to lay down the following three axioms for the theory of properties:

(λ) $\lambda u.\phi(u) = \lambda u.\psi(u) \leftrightarrow Eqv_u(\phi(u), \psi(u))$

(P) $P(\lambda u.\phi(u))$

(η) $v\ \eta\ \lambda u.\phi(u) \leftrightarrow \phi(v)$

Here (λ) gives the identity criteria for properties, which are the same as for concepts, (P) introduces a predicate true of all properties, and (η) the relation of property possession.

Linnebo's main concern in the paper is mathematical concepts, which are most naturally regarded as extensional, and hence he suggests the following simplification of (λ):

(V) $\lambda u.\phi(u) = \lambda u.\psi(u) \leftrightarrow \forall u(\phi(u) \leftrightarrow \psi(u))$

which is, as Linnebo makes clear, just a variant of Frege's Basic Law V.[36] Thus some restriction on what properties there are is needed in order to avoid contradiction.

Suppose that such a restriction has been given. Linnebo's idea is then to let objects and properties (which are just a special kind of objects) be the semantic values of non-logical constants of first-order languages. More precisely, given that $\mathscr{L}$ is a first-order language, Linnebo understands by a *lexicon of* $\mathscr{L}$ a set-theoretic function that maps singular $\mathscr{L}$-terms to (singular) objects and $\mathscr{L}$-predicates to properties. A lexicon may then be expanded

[36] See page 28.

to an *interpretation* by means of set-theoretic recursion on the complexity of formulas.

Semantic theorising involving interpretations of the sort described takes place in a theory T formulated in the language of ZFCU with identity, expanded with the language of the theory of properties. We follow Linnebo and add 1 as a subscript for this language $\mathscr{L}_1$. The $\mathscr{L}_1$-theory T is now said to be *minimally adequate* if it contains enough set theory to handle *n*-tuples and recursion on the syntax of at least first-order languages together with what Linnebo calls the *minimal theory of properties*, or V^- for short.[37] V^- contains an axiom for the existence of the identity property, the following 'part' of (V),

$$\text{(V=)}\quad Px \land Py \to (x = y \leftrightarrow \forall u(u\ \eta\ x \leftrightarrow u\ \eta\ y)),$$

and a number of axioms that close the property of being a property under certain *basic operations*. These operations handle the algebraic operations corresponding to the logical constants of the object language when applied to properties. Thus, for example, in order to handle negation, we need an axiom saying that if p is a property, then so is the complement of p. Likewise, we need to secure the existence of the conjunction of two properties, permutation of *n*-tuples, projecting *n*-tuples, etc.[38]

V^- does not prove the existence of all the properties we would like to have in a semantic theory T. For instance, it does not prove the existence of the property of being an interpretation. It is true, as Linnebo points out, that given a Gödel numbering of T, we may construct a syntactic property, that will take the form of a number theoretic formula, of being an interpretation in T. As usual we may proceed to prove results about the arithmetised theory. Without further axioms in T, however, there is no way to prove the existence of the proper property of being an interpretation, i.e., the existence of an object i such that for all x, $x\ \eta\ i$ if and only if x is an interpretation.

Besides the property of being an interpretation it is natural to ask for interpretations that interpret certain object languages in their intended ways.

[37] Linnebo (2006, p. 164).

[38] Linnebo (2006, p. 163). See also Holmes (1998, ch. 6) where a similar idea is worked out in detail with NFU plus pairing playing the role of T.

For instance, Linnebo points out that we might want to add a property comprehension axiom for the existence of the property of being an ordered pair $\langle x, y \rangle$ such that $x \in y$. That is an instance of

$$(\mathrm{V}\exists) \quad \exists x(Px \wedge \forall u(u \, \eta \, x \leftrightarrow \phi(u))),$$

where $\phi(u)$ expresses that u is an ordered pair where the first component is an element of the second component.

Not all instances of (V∃) give rise to properties however. For instance, it is immediate that '$\neg u \;\; \eta \;\; u$' cannot be allowed since it gives rise to a contradiction in familiar ways. Likewise, the property used in WA, i.e., the property of not being an interpretation such that a predicate letter 'P' is true of that interpretation under that very interpretation, must somehow be barred from being an instance. Thus, the theory Linnebo proposes needs to be supplemented with a restrictive theory of properties.

> To be acceptable, a proposed restriction on the property comprehension scheme must satisfy two potentially conflicting requirements. Firstly, the restriction must be liberal enough to allow the properties we need in order to carry out the desired semantic theorizing. Secondly, the restriction must be well motivated. As an absolute minimum, the restriction must give rise to a consistent theory. But ideally, the restriction should be a natural one, given an adequate understanding of properties. The restriction should be one it would have been natural to impose anyway, even disregarding the fact that paradox would otherwise ensue. (Linnebo, 2006, p. 166)

In a note Linnebo points out that the idea of augmenting ZFCU with a theory of properties to enable semantic theorising is compatible with a range of different understandings of properties. The particular suggestion he advocates involves a restriction on what properties there are that is based on a requirement that the individuation of properties needs to be well-founded. This means that the *fundamental specifications* of an object and the equivalence relation, or *unity relation* that settles when two such specifications determine the same object, are well-founded in the sense that they must not involve, or presuppose any object not yet individuated. In particular, the individuating of an object must not involve, or presuppose, the object itself.

Linnebo makes this mathematically precise by saying that a condition $\phi(u)$ presupposes only those objects that have already been individuated, and no others, if it is invariant under permutations π that fix all objects that have already been individuated and respect all relations that have been individuated in the sense that

$$\forall x_0, \ldots x_n(R(x_0, \ldots, x_n) \rightarrow R(\pi(x_0), \ldots, \pi(x_n))).$$

Such conditions, i.e., conditions presupposing only already individuated objects, define properties. The philosophical motivation behind this is that if properties are given by such conditions they are specified in a non-circular way. One effect of the requirement is that a condition determines a property (by nominalising the concept defined by the condition) only if it does not distinguish between objects that have not yet been individuated. If a condition distinguishes between objects not yet individuated, the condition does not determine a property.

In particular, Linnebo points out, this means that conditions that involves η are problematic. For in its intended sense, η distinguish every property from every other, by means of (V=), and thus "maximally violating the well-foundedness constraint" (Ibid., p. 171). This makes Linnebo suggest that property comprehension for the condition '$u\ \eta\ x$' may come with a restriction that x must be a property already individuated. Then the condition cannot distinguish between objects not yet individuated and, hence, satisfies the well-fondedness criterion. A drawback is that the intended property of η is no longer available and, thus, that semantic pessimism threatens. Linnebo solves the lack of a real η-property by suggesting a hierarchy of more and more inclusive properties interpreting 'η'. These properties give rise to a hierarchy of theories, each of which allows the definition of an interpretation for its predecessor. He concludes:

> Many people will no doubt feel this as a loss. But it should be kept in mind that the rejection of such a property is not an *ad hoc* trick to avoid paradox but follows from the independently motivated well-foundedness requirement. So perhaps

> we must give up as illusory our apparent grasp of an absolutely general relation of property application. (Ibid., p. 175, italics in the original)

In light of the conclusions of Chapter 4, Linnebo's approach is indeed appealing. However, some further points need to be addressed. Linnebo himself points to the need of further investigations into just how unnatural the lack of an absolutely general relation of property application is. Also, just how natural the well-foundedness requirement is needs to be investigated, as well as if it really delimits the properties appropriately.

5.5 Concluding remarks

The type-theoretic semantic theories and Linnebo's proposal with sets and properties are all based on the idea that what is needed, or wanted, is a strictly adequate, or completely general semantics. Otherwise semantic pessimism threatens. This requirement entails the idea of a semantic theory that is maximally liberal as to the possible semantic values of the non-logical vocabulary of the object language. In case of type-theoretic semantics this leads to certain expressive limitations which may be taken to challenge those theories on the very point of being strictly adequate. If we use some concept in the metalanguage, this concept should also be available in the object language as a possible interpretation of an appropriate predicate. We saw that the concept of strict adequacy itself is not so available, nor are other examples that are both natural and important for the type-theorist.

In the next chapter we develop a semantic theory which is inspired by Linnebo's, in many ways appealing, approach. The main differences result from the diverse analyses of what lessons should be drawn from WA. While Linnebo acknowledges the need to provide a theory of properties that shows why the alleged property designated by '*R*' is illegitimate, we argued in Chapter 4 that WA itself is sufficient to that end. Thus, without formulating a separate theory of properties, we develop a model-theoretic semantics very much in the standard way using a non-standard set theory as our metalanguage.

6 Model Theory with a Universal Domain

Model-theoretic semantics of both formal and natural languages has generally been very successful. Of particular interest is that it has proven its usefulness in the semantic analysis of quantifier expressions, again both in formal and natural languages.[1] However, as pointed out many times by now, model-theoretic semantics as formulated within standard set theory is not sufficient for our purposes. The lack of a universal set simply makes it impossible to adequately represent the universal domain of quantification. But due to the general success of the model-theoretic approach, and the arbitrariness of the fact that the metatheory we are accustomed to—standard set theory—happens to lack the means to represent the universal domain, it is well motivated to investigate the model-theoretic approach within some alternative metatheory. That is, it is well motivated to consider the idea of constructing a model-theoretic semantics for absolute quantification within a set theory with a universal set.

Such an approach can be met with two types of critique. First, one may criticise the very idea of providing a model-theoretic semantics for absolute quantification, and, secondly, one may criticise the adopted metatheory in which such an approach is developed. In Section 6.1 critique of both kinds is discussed. We then briefly introduce $\mathrm{NFU_p}$, our metatheory, and develop a model-theoretic semantics in Section 6.2. We take care that our defined concepts, e.g., *model* and the relation of *satisfaction*, correspond to sets. Thus we are able to show that there is a model in our semantics for a first-order formulation of the semantic theory we construct. A completeness theorem for first-order logic in the resulting semantics is also given.

[1] Indeed, Peters and Westerståhl (2006) employ a model-theoretic framework in *Quantifiers in Language and Logic*, which up to date is the most comprehensive treatment of the semantics of quantification.

6.1 Two lines of critique

6.1.1 Williamson's argument and the notion of strictly adequate model-theoretic semantics

Williamson's argument, WA, which we discussed in Chapter 4, is often celebrated as one of the most general arguments against absolute quantification. General, that is, in the sense that it makes minimal presuppositions about the concepts it employs. Thus the argument may be adapted to suit any suggested semantics that accepts its premisses: simply use the concepts of the suggested semantics and run through each step of the argument. In particular, if WA is general in this way, one may think it should be possible to apply it as an argument against any *model-theoretic* semantics of absolute quantification as developed within a set theory with a universal set. In fact, this is precisely what Rayo and Williamson (2003) do. Although we have already explained why WA is inconclusive as an argument against absolute quantification, it is nevertheless instructive to dwell upon some parts of Rayo and Williamson's formulation of this particular instance of the argument.

Having concluded that model-theoretic semantic as formulated within ZF is insufficient due to the lack of a universal set V, and that imposing V as a proper class is of no help since it will not contain itself and thereby fail to be truly universal, Rayo and Williamson proceed to comment on the possibility of using set-theories with a universal set:

> Unfortunately, set theories that allow for a universal set must impose restrictions on the axiom of separation to avoid paradox. So, as long as an MT-interpretation [model-theoretic interpretation] assigns a subset of its domain as the interpretation of a monadic predicate, some intuitive interpretations for monadic predicates will not be realized by any MT-interpretation. (Rayo and Williamson, 2003, p. 333)

Clearly, the axiom of separation

(1) $\forall x \exists y \forall z (z \in y \leftrightarrow z \in x \land \varphi)$, y not free in φ

must fail in any set theory with a universal set. Otherwise we could regain naive comprehension

(2) $\exists y \forall z(z \in y \leftrightarrow \varphi)$, y not free in φ

by putting V for x to get

(3) $\exists y \forall z(z \in y \leftrightarrow z \in V \wedge \varphi)$

Since naive comprehension is inconsistent one must either give up V as a set or the axiom of separation. One way to give up the axiom of separation is to impose restrictions on it: hence the first sentence of the quote.

However, thus explained, the quote may be understood to indicate that each of the founders of a set theory with a universal set, at a certain point, faced the situation we reached through (1) and (2), i.e. that they had to choose between restricting the axiom of separation or giving up V as a set. But this is at least historically misleading. Quine (1937), for instance, developed NF with an eye to the simplified theory of classes in *Principia Mathematica*, and the problem there was not the lack of a universal class, but rather that there were too many of them, viz. one for each type.[2] Moreover, there is no axiom of separation to be restricted in *Principia Mathematica*, but rather a multiplicity of comprehension schemes, again one for each type. Thus, taking NF as an example, it would be a mistake to understand the first sentence of the quote to implicate that Quine at some point *restricted* the axiom of separation to get V.

Indeed, it is true that we get NF from the schema of naive comprehension if we restrict the permitted substitution instances of φ to the so-called stratified formulas, a subclass of well-formed formulas, and add the axiom of extensionality. But then again, taking the historical point of view, we don't get NF from Z, the set theory of Zermelo (1908), which would have been Quine's natural alternative point of departure besides *Principia Mathematica*, even if we rid it of urelements and explicate the Axiom of *Assonder-*

[2] This is an anachronistic remark since signs for classes are contextually defined in *Principia Mathematica* so that any such expression's purported reference to a class vanishes when the notation is expanded into primitive notation. However, Quine (1937) explicitly refers to the simplified theory of types in *Principia Mathematica* for which the claim holds. See Quine (1941) for a discussion.

ung as the axiom of separation. For Z contains the axiom of choice (AC), and, as shown by Specker (1953), NF $\vdash \neg$AC.

Of course, none of this proves Rayo and Williamson wrong in stating that the set theories with a universal set "must impose restrictions on the axiom of separation to avoid paradox," but it shows that, if not read with care, it may blur interesting conceptual distinctions.

The point made in the quoted passage seems to be that, since separation fails in set theories with a universal set, not all subclasses will be subsets, and thus there will be intuitive interpretations of monadic predicates, i.e., the proper subclasses, that will lack a corresponding subset realising the interpretation. Rayo and Williamson go on to claim that this situation is just an instance of Williamson's general argument from paradox. We have already discussed that argument in Chapter 4 and shall not repeat the details here.

Having derived the contradiction, Rayo and Williamson conclude:

> It follows that there are legitimate assignments of semantic values to [predicates] that cannot be captured by any MT-interpretation. (It is worth noting that, although the argument is structurally similar to standard derivations of Russell's paradox, it does not rest on any assumptions about sets. As long as the variables in the metalanguage range over everything, all we need to get the problem going is the observation that MT-interpretations are individuals, and the claim that, whatever it is to G, there are legitimate assignments of semantic values to variables according to which the predicate-letter P applies to something if and only if it Gs.) (Rayo and Williamson, 2003, p. 334).

The last sentence of the quote may be equated with either the substitutional or the semantic version of strict adequacy.[3] Thus, given a set theory T with a universal set, it may be formalised, using the terminology from Chapter 4, in two different ways:

(4) $\exists y(INT(y) \wedge \forall x((y \models P[x]) \leftrightarrow \phi(x)))$ holds for all ϕ

[3] See page 118.

which corresponds to the substitutional version of strict adequacy, and

(5) $\forall z \exists y(INT(y) \wedge \forall x((y \models P[x]) \leftrightarrow x \in z))$

which corresponds to the semantic version.

Let's take seriously that the semantic values of the non-logical vocabulary are sets of some set theory T. Then, under the semantic version of strict adequacy, the claim that there are legitimate assignments of semantic values that cannot be captured by any model means that there are sets that cannot be the semantic values of predicates according to any model. But $x \not\models P[x]$, which is the predicate used to derive the contradiction in WA, determines a semantic value in T only under the additional assumption that

(6) $\mathrm{T} \vdash \exists y \forall x(x \in y \leftrightarrow (x \not\models P[x]))$

However, as we shall see, in NF, or any of its descendants, $x \not\models P[x]$ need not determine a set. Consequently, WA, interpreted with the semantic version of strict adequacy as one of its premisses, is inapplicable to such theories, contrary to the conclusion in the quote that there are legitimate assignments of semantic values that cannot be captured by any model-theoretic interpretation.

Alternatively we may understand Rayo and Williamson as claiming that, just as there are formulas in T that fail to determine sets, there are formulas in the language of any model-theoretic semantics such that their intuitive semantic values cannot be equated as real semantic values in the adopted semantic theory. Note that this is a problem only if we have an additional argument that such formulas should determine a semantic value. But, as we saw in Chapter 4, there is no conclusive argument that $x \not\models P[x]$ should define a set, or, more generally, signify a semantic value. On the contrary, we concluded that what WA shows is that $x \not\models P[x]$ should *not* be considered a legitimate substitution instance of (4).

It is tempting at this point to ask for a characterisation of what formulas count as legitimate instances of (4). A natural suggestion would be to use the notion of semantic value. An $\mathscr{L}_T$-formula φ is then legitimate if

$$\mathrm{T} \vdash \exists x \forall y(y \in x \leftrightarrow \varphi)$$

But this suggestion might turn out to be problematic. Assume that we adopt some metatheory T and start to construct our model-theoretic semantics by defining semantic concepts such as *satisfy*, *model*,..., etc. If it turns out that some of these semantic concepts cannot have their intended meanings according to the semantic theory one might argue that the theory fails to be sensitive to the idea behind substitutional strict adequacy. In case $x \not\models P[x]$ we have a separate argument to the effect that it should not be a legitimate substitution instance to (4), but as long as we have no such arguments for the concepts we define, they should not be ruled out beforehand.

At the same time, as WA shows, we cannot adopt (4) as it stands. What we do construct in this chapter is a model-theoretic semantics for which the following holds for all formulas ϕ that define semantic concepts in our semantic theory:

(7) $\exists y(INT(y) \wedge \forall x((y \models P[x]) \leftrightarrow \phi(x)))$

6.1.2 Metatheories with a universal set

In this section we discuss some critical remarks that have been made concerning the possible meta-theories for a model-theoretic semantics with a universal domain.

Despite the naturalness of the idea, the employment of a set theory with a universal set is generally rather swiftly rejected in the literature. The most common counterarguments to the employment of such set-theories have been nicely summed up by Linnebo (2006, p. 156):

> [...] all known set theories with a universal set, such as Quine's New Foundations, are not only technically unappealing but have lacked any satisfactory intuitive model or conception of the entities in question. It would therefore be folly to trade traditional ZFC for one of those alternative theories.

Since trading ZFC for an alternative set theory is precisely what we do later in this chapter, we will have to address each of the points made by Linnebo in order to see whether they constitute valid arguments against the idea of set-theories with a universal set.

Consider the statement that set-theories with a universal set are technically unappealing. I am not sure exactly what Linnebo has in mind here, and different theories may be considered technically unappealing for different reasons. However, one common feature of all such theories is that none of them has full $\in$-induction; they cannot since $\in$ is not well-founded, allowing e.g. $V \in V$. This may certainly be seen as a technical inconvenience that makes such theories unappealing.[4]

On the other hand, the lack of full $\in$-induction in the metatheory is the kind of inconvenience that one would expect in any semantic theory with the universal domain; such a domain must be a member of itself in order to be truly universal, and, as a consequence, the metatheory cannot have a well-founded membership relation. There seems to be no a priori reason to require that a set theory adopted as metatheory should have full $\in$-induction. Thus, although the lack of full $\in$-induction may be considered an inconvenience, it is hardly something we can count as an argument against using a set theory with the universal set in the present context. In fact, the need of a truly universal domain in the semantic theory, and thus a representation thereof in the metatheory, indicates that, to a certain degree, one should expect that any suitable metatheory has to be technically inconvenient to some extent.

Also, just as different theories may be considered technically unappealing because they have, or lack, certain properties, a theory may also gain or lose its technical attractiveness with regard to its utility, which again is a relative notion; relative, that is, to the kind of work one asks of the theory.

Let us return to Zermelo's theory Z in this context. One aspect of that theory is that only relative complements of sets are available. Otherwise, if A is a set and its complement $A^c = \{ x \mid x \notin A \}$ is also a set, so is V by the union axiom. But V cannot exist as a set because of *Assonderung* and Russell's paradox.

For Zermelo, this is not something which makes his theory unappealing,

[4] For instance, as Forster (1995) notes, ill-founded set-theories have no (transfinite) recursive method of deciding identity between sets.

since the lack of (real) complements is not an obstacle for what he sets out to be his project:

> Set theory is that branch of mathematics whose task is to investigate mathematically the fundamental notions "number", "order", and "function", taking them in their pristine, simple form, and to develop thereby the logical foundations of all of arithmetic and analysis; thus it constitutes an indispensable component of the science of mathematics. (Zermelo, 1908, p. 200)

But there are other ideas about the task of set theory. We, for instance, want to say that set theory is that branch of mathematics which permits us to investigate, with mathematical precision, fundamental *semantic* notions such as truth and logical consequence. This does not rule out that we also want our set theory to be the branch of mathematics in which the mathematical notions Zermelo mentions are investigated. Still, it provides a point of view from which the presence of non-relative complements, as well as the presence of V, may be considered technically appealing.

Linnebo's remark that it would be folly to trade ZFC for some alternative set theory because such theories are technically unappealing is hardly meant as a conclusive argument. But even as a critical remark, it stands in need of further elaboration. Our short discussion seems to show that the attribute *technically unappealing* is relative to, among other things, the job the set theory is supposed to do. In light of that it is perhaps better to postpone the estimation of the technical attractiveness of the theory until it has proven its utility.

At this point one could argue that the lack of a universal set, or the lack of real complements, is not really a technical aspect of a theory, but typically results from one's conception of sets. Thus we are led to Linnebo's second and third lines of critique, that set theories with a universal set "have lacked any satisfactory intuitive model or conception of the entities in question."

This should be contrasted with the situation in ZFC where the iterative conception of sets is commonly regarded as giving rise to the cumulative universe of sets which, in turn, standardly serves as its intuitive model. Furthermore, as (Boolos, 1971, p. 16) puts it: "*the iterative conception of set*

[...] often strikes people as entirely natural, free from artificiality, not at all ad hoc, and one they might perhaps have formulated themselves."

The iterative conception of sets is often taken to be so closely related to the cumulative hierarchy that one does not bother to tell them apart. But, as Forster (2008) demonstrates, an iterative conception of sets may give rise to structures other than the cumulative hierarchy. Indeed, Forster's construction gives rise to a model of "roughly" the set theory of Church (1974) which, incidentally, has V as a set.[5] This may be taken to show that the iterative conception is less sharp than usually assumed and that there are, in fact, several possible iterative conceptions of sets. Hence the following disclaimer: in the sequel, 'the iterative conception of sets' is taken to refer to that conception which gives rise to the cumulative hierarchy.

It is not known if Zermelo had the iterative conception of sets in mind when axiomatising Z. What seems to be clear, however, is that, whatever conception of sets Quine saw as underlying the axioms of Z, he considered it unsatisfactory when working on NF. In a brief note circulated at the colloquium on NF's 50th birthday, in Oberwolfach 1987, he comments in retrospect on the motivations behind his theory. Having explained two drawbacks he saw in the simplified version of the set theory of *Principia Mathematica*, he continues:

> Zermelo's system itself was free of both drawbacks, but in its multiplicity of axioms it seemed inelegant, artificial, and ad hoc. I had not yet appreciated how naturally his system emerges from the theory of types when we render the types cumulative and describe them by means of general variables. I came to see this only in january 1954, [...]. If I had appreciated it in 1936, I might not have pressed on to "*New Foundations.*"
>
> But I might have still. For I disliked the lack of a universe class in Zermelo's system, and the lack of complements of

[5] Forster suggests that we create both sets and their complements at each stage. Thus, at the first stage, when there is nothing yet created, we get $\emptyset$ and its complement V, at the second stage we get in addition to what we already have $\{\emptyset\}$, $\{V\}$, $V \setminus \{\emptyset\}$, $V \setminus \{V\}$, $V \setminus \{\emptyset, V\}$ and $\{\emptyset, V\}$. And so on through the ordinals.

> classes, and in general the lack of big classes. (Quine, 1995, p. 287–288, italics in the original.)

Thus, whatever conception of classes Quine had back in 1936, it was not the same conception as Zermelo must have had when axiomatising Z: that much is clear from the final sentence of the quote.

As it turned out, Z can in fact be seen as an early axiomatisation of the cumulative hierarchy based on an iterative conception of set. Linnebo's critical remark amounts to asking: if we are to trade standard set theory for some alternative, NF say, what then are the conceptions and structures corresponding to the iterative conception and the cumulative hierarchy? For it seems to be Linnebo's view that, if we are to abandon standard set theory, we should at least look for something equally well motivated in terms of intuitive models and conceptions of sets. If we can't find such a theory, we shouldn't make the trade at all.

Consider now the notion of an intuitive model. The situation seems reasonably clear when it comes to affirming the existence of an intuitive model, e.g., of standard set theory, but claiming that alternative set theories lack intuitive models is less straightforward. For we have no precise definition of *intuitive model* and without a definition it is hard to make the claim that there are no intuitive models for the theories in question.

To some extent the situation parallels the situation before Turing's and Church's analyses of the concept of *decision procedure* in the thirties. To show that a problem is decidable amounts to providing some decision procedure, e.g. an algorithm, for its solution. In the very same way, to claim that a set theory has an intuitive model, amounts to giving such a model. On the other hand, in order to claim that some problem is undecidable one needs to prove that there is no decision procedure for that problem, and, of course, this cannot be accomplished unless we have a precise definition of what a decision procedure is. In the same way, the statement that some set theory lacks an intuitive model calls for a precise definition of what an intuitive model would be.

Now, we do have an exact definition of the concept of algorithm, e.g., in terms of Turing machines, and are thus entitled to claim that certain problems are undecidable via the identification of decision procedures and

Turing machines. However, the situation is less clear when we turn to intuitive models of set theories. One problem has to do with circularity. We are accustomed to talk about models in terms of certain sets, and it is hard to imagine an explication of *intuitive model* that is very different from our ordinary models. Another problem is that the concept of *set* is a basic concept. It is hard to see what other basic concepts we may use to explicate the notion of intuitive model. Thus, we may conclude, until we have a clear conception of what the alternative set theories lack, i.e., until we have a clear conception of what an intuitive model is, it seems premature to rule out all such theories as possible metatheories.

Linnebo does not really claim, in the quoted passage, that all alternative set theories hitherto presented lack intuitive models of the right kind. Rather he says that, so far no satisfactory intuitive models have been presented. This is of course consistent with there being undiscovered intuitive models.

In any case, the situation calls for further analysis of the consequences of using a set theory lacking intuitive models. Dummett suggests that the lack of a model, intuitive or otherwise, equals the lack of a subject-matter. Interestingly he claims this with an eye to NF:

> Whatever mathematicians profess, mathematical theories conceived in a wholly formalist spirit are rare. One such is Quine's New Foundation system of set theory, devised with no model in mind, but on the hunch that a purely formal restriction on the comprehension axiom would block all contradictions. The result is not a mathematical theory, but a formal system capable of serving as an *object* of mathematical investigation: without some conception of what we are talking about, we do not have a theory, because we do not have a subject-matter. [...] if an angel from heaven assured us of its consistency, we should still not have a mathematical theory until we attained a grasp of the structure of a model for it. (Dummett, 1991, p. 230).

Suppose we trade standard set theory for NF. If Dummett is right we would find ourselves working in a theory destitute of a subject-matter; it would not be *about* anything in the sense that Peano arithmetic, or ZFC, is about the structure of natural numbers, or the cumulative hierarchy of sets, respectively. An awkward position indeed. On the other hand, if we instead trade standard set theory for NFU_p, i.e., the theory that results from NF by adding urelements and a pairing operator, the situation is slightly different. First, NFU_p is provably consistent relative to ZFC and, second, even though there is no intuitive model available, we do have some grasp of some of its models.

But even with no model in mind, it may be still be possible to argue that we can use NF-like theories in a non-formalistic way, i.e. as theories about sets. NF belongs to the logicist tradition where classes are considered to be properties in extension.[6] Thus conceived it is quite natural that $\in$, now a relation between individuals and extensions, isn't well founded. For there are doubtlessly extensions that belong to themselves; the extension of 'extension' being one example. Also, for any property, there is the property of not having that property, so full complements are natural under such a

[6]To see this remember that type theory, as presented in Russell (1908) and in Russell and Whitehead (1910), is a ramified hierarchy of propositional functions which are categorised according to (i) what kind of, and how many, arguments they take, and (ii) what bound variables they contain. Thus, for instance, a propositional function of one variable taking individuals as values will be of the first order if it involves no bound variables except variables ranging over individuals; it will be of the second order if it contains only bound individual variables and at least one variable ranging over first-order functions of one variable, and so on. The hierarchy of orders provides an intensional character to the propositional functions, for two equivalent propositional functions, i.e. two functions true of the same arguments, may belong to distinct orders and thereby fail to satisfy the same propositional functions.

Also, consider the following instructive passages:

> A propositional function of *x* may, as we have seen, be of any order; hence any statement about 'all properties of *x*' is meaningless. (A 'property of *x*' is the same thing as a 'propositional function which holds of *x*'.)
>
> […]
>
> Hence, we must find, if possible, some method of reducing the order of a propositional function without affecting the truth or falsehood of its values [i.e. propositions]. This seems to be what common sense effects by the admission of *classes*.(Russell, 1908, pp. 241–242)

Thus, it is clear that Russell understood classes as properties in extension.

conception. Likewise is it natural to assume an extension, V, of the property of being self-identical.[7]

Thus, even though we may have no clear understanding of what the universe of NF-sets may be like, there is still room, via our intuitions of properties in extension, for non-formalistic reasoning about such sets.

6.2 A model-theoretic semantics for absolute quantification

In this section we carry out our plan of using a set theory with a universal set and construct a model-theoretic semantics for first-order languages in NFU_p. This theory is introduced in Section 6.2.1. The definition of the model-theoretic semantics is given in Section 6.2.2, and a completeness result is given in Section 6.2.3.

6.2.1 NFU_p

NFU_p is a classical first-order theory with the membership relation, $\in$, and an operator, $\langle\ ,\ \rangle$, taking objects to ordered pairs.[8]

NFU_p shares with NF its strategy to avoid the set-theoretical paradoxes. The idea is to circumscribe the naive axiom schema of comprehension, $\exists y \forall x (x \in y \leftrightarrow \varphi(x))$, by restricting the class of formulas instantiating it to the so-called *stratified* formulas.

Definition 6.2.1. A formula φ in $\mathscr{L}_{\text{NFU}_\text{p}}$ is *stratified* if there is a function τ from the (not necessarily free) variables in φ to $\mathbb{N}$ such that

- if '$x = y$' or '$\langle x, y \rangle$' occurs in φ, then $\tau(x) = \tau(y)$;
- if '$x \in y$' occurs in φ, then $\tau(y) = \tau(x) + 1$'.

τ is a *stratification* if it fulfils the conditions above.

Thus, for example, '$x \notin x$' and '$x = y \land z \in x \land y \in z$' are unstratified, whereas '$x \notin y$' and '$x = y \land x \in z \land y \in z$' are stratified. Given a formula

[7] Informal discussions on these matters have appeared on the "FOM-list". For a published formulation, see Forster (1995).

[8] The standard reference for NFU_p-studies is Holmes (1998).

φ, the value of an associated stratification τ for a variable x in φ is called the *type of x in φ under τ*. Most often we just talk about the *type of x*, assuming the rest to be clear from the context.

The axiom schema of naive comprehension is now restricted to stratified formulas:

Axiom 6.2.2 (Stratified comprehension)**.** $\exists x \forall y (y \in x \leftrightarrow \varphi(y))$, where φ is a stratified formula in which x is not free.

We use $\{ x \mid \varphi \}$ to denote a set whenever φ is a stratified formula. Note, also, that even if φ has no stratification, it may be equivalent to a stratified formula, so that it is understood to specify a set $\{ x \mid \varphi \}$ in an indirect way. Note also that, though stratified comprehension gives us the means to prove the existence of a great number of sets, it provides no direct method of proving the non-existence of a certain alleged set. Stratification is a sufficient, but not necessary, condition for sethood. Thus, even if the existence of a set $\{ x \mid x \notin x \}$ does not follow from stratified comprehension, since $x \notin x$ is not stratified, we need Russell's argument, i.e. that the existence of such a set leads to contradiction, to conclude that it cannot exist.

If sets are perceived as extensions of predicates, the requirement of stratification thus induces a principle of precaution: any extension defined in a stratified way exists, and if there is no stratification for an alleged extension, accept it as existing only if there is no strong argument to the contrary.

It follows from the axiom of stratified comprehension that there is a universal set $V = \{ x \mid x = x \}$, and for any set y, its complement $y^c = \{ x \mid x \notin y \}$ exists.

The axiom of stratified comprehension together with the axiom of extensionality constitutes the set-theoretic part of NF. We could choose to adopt NF as our metatheory, but then we would face at least two problems. First, as was shown by Specker (1953), NF proves the negation of the axiom of choice, which makes it an unfriendly theory to work in. Second, and more importantly, the consistency of NF remains an open problem. To argue that NF is a suitable metatheory would be to beg the question on its consistency.

Also, though Quine (1937, pp. 81–82) understood the variables of NF as ranging over absolutely everything, including urelements, or non-sets, the axiom of extensionality in NF seems to imply that there can be at most one urelement; urelements have no elements, so, under extensionality, they are all identical. Quine solved this problem by stipulating that $x \in y \leftrightarrow x = y$ whenever y is an urelement. If extensionality holds, this has as a consequence that, if x is a urelement, $x = \{x\}$; for, clearly, $x \in x \leftrightarrow x \in \{x\}$ for all urelements x. Thus, since $\{x\} \neq \{y\}$ for $x \neq y$, urelements are differentiated in a roundabout way by extensionality. But even if this solution has the right consequences, it is nevertheless somewhat ad hoc.

Another way of imposing urelements in NF, suggested by Jensen (1968-69), is to weaken the axiom of extensionality by restricting it to non-empty sets:

Axiom 6.2.3 (Weak extensionality)**.**

$\forall x \forall y (\exists z (z \in x) \wedge \forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y)$

Weak extensionality together with stratified comprehension constitutes NFU, which in many respects is quite different from NF. Jensen (1968-69) proved that NFU + AC + Inf is consistent relative to ZFC.[9] Thus, the problems that made us avoid the adoption of NF as our metatheory are not relevant as arguments against the adoption of NFU.

The binary operator '$\langle\ ,\ \rangle$' was not part of NFU as originally formulated by Jensen (1968-69), but is added for convenience. It is governed by

Axiom 6.2.4. $\forall x, y, z, w (\langle x, y\rangle = \langle z, w\rangle \rightarrow x = z \wedge y = w)$

In Definition 6.2.1 we stipulated that, if a formula contains an ordered pair $\langle x, y\rangle$, the formula cannot be stratified unless the type of x is also the type of y. We did not assign a type to the ordered pair itself—only variables are assigned types—but we need some rule governing how to assign a type to z in a stratified formula containing $z = \langle x, y\rangle$. It is convenient to let pairs have the same type as their objects. I.e., if $z = \langle x, y\rangle$ appears in a stratified formula with a stratification τ, then $\tau(z) = \tau(x) = \tau(y)$.

$\mathrm{NFU_p}$ is NFU with the ordered pair operator $\langle\ ,\ \rangle$, and Axiom 6.2.4.

[9] In fact, Jensen showed that NFU + AC + Inf is consistent relative to the simple theory of (finite) types with AC and Inf.

We construct complex names by means of definite descriptions in, more or less, the standard way

$$\psi[(\imath x)(\varphi)] \Leftrightarrow (\exists! x\varphi \rightarrow \forall x(\varphi \rightarrow \psi)) \wedge (\neg\exists! x\varphi \rightarrow \forall x(x = \emptyset \rightarrow \psi))$$

Thus, if the condition φ of the description is satisfied by nothing, or by more than one object, the definite description denotes the empty set.

Once we have ordered pairs, finite sequences may be inductively defined, and n-ary relations may in turn be defined as sets of ordered n-tuples (i.e. sequences with n terms), and n-ary functions as sets of ordered $n+1$-tuples in the standard way. If a unary function f is defined as a set of ordered pairs $\langle x, y\rangle$, the type of f, in a stratified sentence, will be one higher than the type of x and y; especially $f(x)$ is of one type lower than f.

We define π_1 as the projection function mapping $\langle x, y\rangle \mapsto x$. Thus, in a stratified sentence, $\pi_1(x)$, if defined for x, will have the same type as x. The same holds for π_2, which is defined as the map $\langle x, y\rangle \mapsto y$. Given an n-sequence $y = \langle y_1, y_2, \ldots y_n\rangle$, the function $proj(i, y) = y_i$ is definable in a stratified way.[10]

It is worth noting that Peano Arithmetic (PA) is interpretable in $\mathrm{NFU_p}$. The natural numbers are defined in Frege-Russell style as equivalence classes on V under equipotence.[11] 0 is defined as $\{\emptyset\}$. The successor (Sc) of a set X of sets is defined as $\{x \cup \{y\} | x \in X \wedge y \notin x\}$. Thus, for instance, the successor of $3 = \{\, x \mid x \text{ has exactly three members} \,\}$ will be the set $4 = \{\, x \mid x \text{ has exactly four members} \,\}$. The set $\mathbb{N}$ of natural numbers is then defined as the intersection of all inductive classes, i.e. classes Y such that $0 \in Y \wedge \forall x(x \in Y \rightarrow Sc(x) \in Y)$.[12] The definitions of addition and multiplication present no difficulties.

[10] See Holmes (1998, p. 40). A finite sequence $\langle y_1, y_2, \ldots y_n\rangle$ is defined as usual as $\langle y_1, \langle y_2, \langle \ldots \langle y_{n-1}, y_n\rangle \ldots\rangle$. $proj(i, y)$ is defined as $\pi_1(y)$ for $i = 1$, as $\pi_2^{i-1}|\pi_1(y)$ for $1 < i < n$, and as π_2^{n-1} for $i = n$. Here $R|Q$ is the relative product of R and Q, R^n is the nth power of R.

[11] Two sets are equipotent if there is a bijection between them.

[12] The existence of a set of the inductive classes follows immediately from stratified comprehension, and so does the intersection over any set of sets.

One worry here is that, since NFU_p contains no explicit axiom of infinity, we may find ourselves in the situation that there is a greatest natural number. However, Rosser (1952) showed that the existence of ordered pairs with the same type as their constituents implies the axiom of infinity.

The interpretability of PA in NFU_p secures that we may use inductive definitions. That is, if φ holds for 0 and, if it holds for the natural number n, then it holds for $n+1$, it follows that φ holds for all natural numbers. The only requisite is that φ is stratified. In fact, since all formulas in the language of PA are stratified, interpretability here implies that NFU_p is a conservative extension of PA, which in turn warrants the use of formulas such as "x is a proof of φ", in inductive proofs. We may also use recursion to define a function F from a function G on natural numbers by means of $G(n, F(n)) = F(n+1)$ if the arguments and the values are given the same type in a stratification.[13] This provides the tools we need in the next two sections, which introduce standard model-theoretic concepts and prove some results about them.

6.2.2 Definitions of semantic concepts in NFU_p

In this section we present a new semantics for absolute quantification by constructing a model-theoretic semantics for first-order languages in NFU_p. As usual we arrive at a definition of truth in a model via satisfaction and valuation. We make sure that all semantic concepts are given stratified definitions. In that way their extensions in NFU_p can be used to define models and the intended extensions of the semantic concepts in the metalanguage become available in interpretations of the object language.

Though the study of a stratified satisfaction relation is a novel approach to the study of absolute quantification, the idea of using a standard satisfaction relation between formulas and structures in NFU_p is not new. Solomon Feferman (1974, 2011) uses NFU_p to provide a foundation for category theory and the general theory of structures.[14] A first-order structure, A, is defined as a tuple of the following form

[13] See Holmes (1998, Chapter 12).

[14] Actually Feferman starts in NFU_p, but eventually shows that an extension of that theory is better suited for his project.

$$A = \langle M, R_1, \dots, R_j, F_1, \dots, F_k, K_1, \dots, K_l \rangle$$

where each R_i is a relation on M, each F_i a function on M, and each K_i a unit set $\{m_i\}$ of some element in M. Given a satisfaction relation, $\models$, the model class, model(θ), of a formula θ is defined in the usual way as $\{A \mid A \models \theta\}$. Feferman associates with each formula θ a stratified formula θ^* such that model$(\theta) = \{ X \mid \theta^*(X) \}$. Then, for example, Feferman can express that the structure $\langle PO, Sub \rangle$, of all partially orderings, PO, ordered under the substructure relation, Sub, is itself a partial ordering,

$$\langle PO, Sub \rangle \in PO,$$

by defining θ^* in the right way. In that way certain natural statements in the informal general theories of structures may be given a straightforward formal representation. Similarly for certain category-theoretic statements.

However, Feferman makes no efforts to stratify $\models$ and his structures, as defined, are of finite signature. Thus, although Feferman's investigations are of considerable importance, further work needs to be done in order to get what we want.

In certain respects interpretations, as defined here, may be conceived as nominalisations of the second-order interpretations defined in Rayo and Williamson (2003). The underlying idea is to take an interpretation to be an ordered couple of a domain and a relation, where the relation is to hold between an individual constant and the object named by it, and between an n-ary predicate and the n-tuples of which it is true, and similarly for function symbols. Thus, in contrast to standard first-order models, where a unary predicate is interpreted as a subset of the domain of the interpretation, the relation in the interpretation as defined here relates the predicate with each object in the domain of which it is true.

Assume that a generic syntax for first-order languages has been defined in NFU_P. In particular, assume for each n that the sets PRED; n of n-ary predicate symbols, FUNC; n of n-ary function symbols as well as the sets CONS of individual constants and VAR of variables, are defined. For a particular language $\mathscr{L}$, the set of n-ary predicates will then be the intersection $\mathscr{L}_{\text{PRED};n} = \mathscr{L} \cap \text{PRED}; n$. Likewise for the other syntactic categories.

Definition 6.2.5.

$$\begin{aligned}
&\mathrm{INT}_{\mathscr{L}}(\mathcal{M}) \Leftrightarrow \exists M \exists I(\mathcal{M} = \langle M, I\rangle \wedge \exists x(x \in M) \wedge I \subseteq \mathscr{L} \times V \wedge \\
&\quad \forall x \forall y \forall n \in \mathbb{N}(x \in \mathscr{L}_{\text{PRED};n} \wedge \langle x, y\rangle \in I \rightarrow y \in M^n) \wedge \\
&\quad \forall x \forall y(x \in \mathscr{L}_{\text{CONS}} \wedge \langle x, y\rangle \in I \rightarrow y \in M \wedge \forall z(\langle x, z\rangle \in I \rightarrow y = z)) \wedge \\
&\quad \forall x \forall y \forall n \in \mathbb{N}(x \in \mathscr{L}_{\text{FUNC};n} \wedge \langle x, y\rangle \in I \rightarrow y \in M^{n+1} \wedge \\
&\quad \forall z(z \in M^{n+1} \wedge \forall i \leq n(\mathit{proj}(i, y) = \mathit{proj}(i, z)) \wedge \langle x, z\rangle \in I \rightarrow \\
&\qquad\qquad \mathit{proj}(n+1, z) = \mathit{proj}(n+1, y))))
\end{aligned}$$

To stratify this, put $\tau(\mathcal{M}) = \tau(M) = \tau(I) = 1$ and let every other variable in the formula receive type 0. Note that, since a finite sequence has the same type as its elements, $\mathcal{M}$ will receive the same type as M and I in a stratified sentence. Note also that M^n, i.e. the n-time iterated product of M, receives the same type as M.

Since $\mathrm{INT}_{\mathscr{L}}(\mathcal{M})$ has a stratified definition, it follows from stratified comprehension that there is a set $\mathrm{INT}_{\mathscr{L}}$ in $\mathrm{NFU_p}$ of $\mathscr{L}$-interpretations. Below we sometimes drop the subscript $\mathscr{L}$.

Next we define the concept *sequence in* $\mathcal{M}$:

Definition 6.2.6. $\langle x, y\rangle \in \mathrm{SEQ} \Leftrightarrow x \in \mathrm{INT} \wedge y : \mathbb{N} \rightarrow \pi_1(x)$

We freely use the convention of writing M, N, etc. for $\pi_1(\mathcal{M}), \pi_1(\mathcal{N})$, etc., whenever $\mathcal{M}, \mathcal{N}$ are interpretations. Likewise, we use I for the second component in an interpretation $\mathcal{M}$, and if it is not clear from context which I belongs to which interpretation, we index it with the interpretation.

Furthermore, since y in Definition 6.2.6 is a function from the natural numbers to the domain of x, in a stratified sentence $\tau(x) = \tau(y)$. Hence, SEQ is stratified.

Definition 6.2.7. If $\mathrm{INT}_{\mathscr{L}}(\mathcal{M})$ for some language $\mathscr{L}$, then the set of sequences in $\mathcal{M}$, $\mathrm{SEQ}_{\mathcal{M}}$, is defined as the set *co-domain*$(\mathrm{SEQ} \restriction \{\mathcal{M}\})$.

The next step consists in defining a valuation (of terms) under some given $\mathcal{M}$ and $s \in \mathrm{SEQ}_{\mathcal{M}}$:

Definition 6.2.8. Let $\mathrm{INT}_{\mathscr{L}}(\mathcal{M})$ and $s \in \mathrm{SEQ}_{\mathcal{M}}$. Then $x \in \mathrm{VAL}_{\mathcal{M},s}$ if and only if x is an ordered pair $\langle t, v\rangle$, $t \in \mathscr{L}_{\text{TERM}}$ and

1. if t is a variable x_i, then $v = s(i)$

2. if t is a constant c, then $v = (\iota z)(\langle t, z\rangle \in I)$

3. if $t = f^n(t_1, \ldots, t_n)$ then
   $v = (\iota z)(\langle f^n, \langle \mathrm{VAL}_{\mathcal{M},s}(t_1), \ldots \mathrm{VAL}_{\mathcal{M},s}(t_n), z\rangle\rangle \in I)$

where $\mathrm{VAL}_{\mathcal{M},s}(t) = (\iota z)(\langle t, z\rangle \in \mathrm{VAL}_{\mathcal{M},s})$.

This formula is stratified with $\tau(x) = \tau(t) = \tau(v)$, and thus, given an interpretation $\mathcal{M}$ and a sequence s in $\mathcal{M}$, $\mathrm{NFU_p}$ proves the existence of $\mathrm{VAL}_{\mathcal{M},s}$. Furthermore, an easy induction on the complexity of t shows that $\mathrm{VAL}_{\mathcal{M},s}$ is a function. It is also (quite) easy to see that there is a function VAL taking an interpretation and a sequence as arguments and giving a valuation. $\mathrm{VAL}_{\mathcal{M},s}$ is the value of this function for the arguments $\mathcal{M}$ and s.

Definition 6.2.9. The relation $\models$ between an interpretation $\mathcal{M}$, a sequence $s \in \mathrm{SEQ}_{\mathcal{M}}$ and a formula φ, $\mathcal{M}, s \models \varphi$, is inductively defined on the complexity of φ:

1. if φ is $t_i = t_j$: $\mathcal{M}, s \models \varphi \Leftrightarrow \mathrm{VAL}_{\mathcal{M},s}(t_i) = \mathrm{VAL}_{\mathcal{M},s}(t_j)$

2. if φ is $P^n(t_1, \ldots, t_n)$:
   $$\mathcal{M}, s \models \varphi \Leftrightarrow \langle P^n, \langle \mathrm{VAL}_{\mathcal{M},s}(t_1), \ldots, \mathrm{VAL}_{\mathcal{M},s}(t_n)\rangle\rangle \in I$$

3. φ is $\neg\psi$: $\mathcal{M}, s \models \varphi \Leftrightarrow$ not $\mathcal{M}, s \models \psi$

4. φ is $\psi \vee \xi$: $\mathcal{M}, s \models \varphi \Leftrightarrow \mathcal{M}, s \models \psi$ or $\mathcal{M}, s \models \xi$

5. φ is $\exists x_i \psi$: $\mathcal{M}, s \models \varphi \Leftrightarrow \exists m \in \pi_1(\mathcal{M})(\mathcal{M}, s[m/i] \models \psi)$

The definition of $\models$ will be stratified if $\tau(\mathcal{M}) = \tau(s) = \tau(\varphi)$. The only tricky part is $\mathrm{VAL}_{\mathcal{M},s}(t_i)$ which is supposed to have one type below I, which in turn has the same type as $\mathcal{M}$. But since

$$\begin{aligned}\mathrm{VAL}_{\mathcal{M},s}(t_i) &= (\iota v)(\langle t_i, v\rangle \in \mathrm{VAL}_{\mathcal{M},s})\\ &= (\iota v)(\langle t_i, v\rangle \in (\iota x)(\langle \mathcal{M}, s, x\rangle \in \mathrm{VAL})\end{aligned}$$

we see that this is as it should: $\tau(x) = \tau(\mathcal{M}) = \tau(I) = \tau(v) + 1$, where $\tau(v) + 1$ is the type of $\text{VAL}_{\mathcal{M},s}(t_i)$ in a stratified sentence.

We define some standard semantic concepts:

Definition 6.2.10.

1. A sentence φ is true in $\mathcal{M}$, $\mathcal{M} \models \varphi$, if, and only if, there is a $s \in \text{SEQ}_{\mathcal{M}}$ such that $\mathcal{M}, s \models \varphi$.
2. φ is a logical consequence of Γ, $\Gamma \models \varphi$, if, and only if, $\mathcal{M} \models \Gamma$ entails that $\mathcal{M} \models \varphi$.
3. φ is a logical truth, $\models \varphi$, if it is a logical consequence of the empty set.

As usual, '$\models$' is ambiguous between the relation of a formula being satisfied by a sequence in an interpretation and the relation of logical consequence. There is also an intensional difference between '$\models$' as defined in NFU_p, and '$\models$' as defined in ZFC; a difference that supervenes on the difference in the conception of sets in NFU_p and ZFC. A natural question is if there is also an extensional difference as to which sentences come out as logical truths and which are logical consequences of which.

Before we tackle that question in the next section, we close this section with some further facts:

Fact 6.2.11. Let $\mathcal{M}^\Pi = \langle V, I^\Pi \rangle$, where I^Π is a relation between the symbols of our semantic theory, i.e. '$\models$', 'INT', 'SEQ', 'VAL' and their corresponding sets. Then $\mathcal{M}^\Pi \in \text{INT}$.

Proof. Stratified comprehension and inspection of the definition of INT. $\square$

Fact 6.2.12. NFU_p does not prove the existence of a model $\langle V, I^\in \rangle$, where $I^\in$ holds between '$\in$' and pairs $\langle s(i), s(j) \rangle$ if and only if $\text{NFU}_\text{p} \vdash x_i \in x_j[s]$.

Proof. The existence of $I^\in$ would make NFU_p inconsistent by Russell's paradox: the complement of $I^\in[\in] \cap \{\, \langle x, x \rangle \mid x \in V \}$ is the Russell class. $\square$

Fact 6.2.13. $\{\, x \mid x \not\models P[x] \,\}$, where $P \in$ PRED; 1, is not a set in $\mathrm{NFU_p}$.

Proof. Since $\mathrm{NFU_p}$ proves $\exists y(\mathrm{INT}(y) \wedge \forall x(y \models P[x] \leftrightarrow x \in F))$ whenever F is a set, the existence of $\{\, x \mid x \not\models P[x] \,\}$, which has an unstratified membership condition, would make $\mathrm{NFU_p}$ inconsistent. This is essentially Williamson's argument. □

6.2.3 Completeness

One way to prove that the two consequence relations, i.e., the one in $\mathrm{NFU_p}$ and the one in ZFC, are extensionally equivalent is to use soundness and completeness. We know that $\models$ as defined in ZFC is coextensive with $\vdash$, for some appropriate first-order calculus. So, if $\models$ as defined in $\mathrm{NFU_p}$ is coextensive with $\vdash$ as well, the two relations are extensionally indistinguishable.

Now, while the soundness part of this argument is obvious, the completeness part needs an argument. One way to prove completeness is to show, as a lemma, that each consistent theory has a model. We may then use the fact that, if a sentence φ is not deducible from a set of sentences Γ, $\Gamma \cup \neg\varphi$ is consistent. It follows from the lemma that $\Gamma \cup \neg\varphi$ has a model, which implies that φ is not a logical consequence of Γ. Hence, $\Gamma \nvdash \varphi$ implies $\Gamma \not\models \varphi$. Contraposition gives that $\Gamma \models \varphi$ implies $\Gamma \vdash \varphi$.[15]

The standard proof of the lemma involves a function taking individual constants to their equivalence classes (assuming that identity belongs to the language). However, the existence of such a function is problematic in $\mathrm{NFU_p}$ since its (obvious) definition is not stratified. So the needed function may not exist. Hence, to prove the lemma, we cannot simply mimic the standard proof. Moreover, intuitively it is far from obvious that a first-order calculus should be complete with regard to our semantics. After all, some classes that are sets in ZFC fail to be sets in $\mathrm{NFU_p}$, and vice versa, and thus, some models of the one theory have no counterparts in the other.

It is thus an interesting fact that we nevertheless may prove the following lemma:

[15] Since $\mathrm{NFU_p}$ interprets PA, another, more direct, but perhaps less informative, way to argue for completeness goes via the arithmetised Henkin proof.

Lemma 6.2.14. Let T be a consistent theory in a denumerable language $\mathscr{L}_T$. Then NFU_p proves that T has a denumerable model.

Informal proof sketch. As usual, to get a witness complete extension of T we make sure that each existential sentence is witnessed by some possibly new individual constant. But, instead of adding a denumerable set of individual constants and using the class of equivalence classes under provable identity as the domain of quantification, we will use the individual constants themselves as objects in the domain. Thus, whenever we need to add a constant to witness a formula we also add that constant to the domain and make sure that the extended theory does not prove it to be identical to some individual constant already added. Consistency is then proved in relation to the set of sentences saying that each constant in the domain is different from every other constant in the domain. Once that has been accomplished the proof proceeds in a standard way.

Proof. Let $\theta_0(\bar{x}), \theta_1(\bar{y}), \ldots$ be an enumeration of the $\mathscr{L}_T$-formulas such that each formula occurs infinitely many times in a cofinal manner. Let $d_0, d_1, d_2, \ldots$ be an enumeration of infinitely many individual constants not in $\mathscr{L}_T$. We construct a maximally consistent and witness complete theory T_∞ such that $T \subseteq T_\infty$, and a set D_∞ of individual constants not in $\mathscr{L}_T$.

Let

$$T_0 = T, \text{ and}$$

$$D_0 = \{d_0\}, \text{ where } d_0 \notin \mathscr{L}_T$$

Assume that T_k and $D_k = \{d_0, \ldots d_{n_k}\}$ have been constructed and consider $\theta_k(x_0, \ldots, x_{l_k})$, where $x_0, \ldots, x_{l_k}$ are the free variables of θ_k. Substituting the elements in D_k for the variables in θ_k yields ${n_k}^{l_k}$ sentences which we may assume to be ordered in some suitable way:

$$\theta_{k,1}, \theta_{k,2}, \ldots, \theta_{k,{n_k}^{l_k}}$$

Let $D^{\neq}$ be $\{d_x \neq d_y \mid d_x, d_y \in D \wedge x \neq y\}$, put $T_{k,0} = T_k$, $D_{k,0} = D_k$, and define

$$\mathrm{T}'_{k,i} = \begin{cases} \mathrm{T}_{k,i} \cup \{\theta_{k,i}\} & \text{if } \mathrm{T}_{k,i} \cup \{\theta_{k,i}\} \text{ is consistent with } D^{\neq}_{k,i} \\ \mathrm{T}_{k,i} \cup \{\neg\theta_{k,i}\} & \text{otherwise} \end{cases}$$

We then proceed to define $\mathrm{T}_{k,i+1}$.[16]

$$\mathrm{T}_{k,i+1} = \begin{cases} \mathrm{T}'_{k,i} \cup \{\varphi(d)\} & \text{if } \theta_{k,i} \text{ is } \exists y \varphi(y) \text{ and } d \text{ is the smallest individual constant in } D_{k,i} \text{ such that } \mathrm{T}'_{k,i} \cup \{\varphi(d)\} \text{ is consistent with } D^{\neq}_{k,i} \\ \mathrm{T}'_{k,i} \cup \{\varphi(d)\} & \text{if } \theta_{k,i} \text{ is } \exists y \varphi(y), d \text{ is the smallest individual constant not in } D_{k,i} \text{ and } \mathrm{T}'_{k,i} \cup \{\varphi(d')\} \text{ is inconsistent with } D^{\neq}_{k,i} \text{ for all } d' \in D_{k,i} \\ \mathrm{T}'_{k,i} & \text{otherwise} \end{cases}$$

We also put

$$D_{k,i+1} = \begin{cases} D_{k,i} \cup \{d\} & \text{if } d \notin D_{k,i} \text{ is used in the construction of } \mathrm{T}_{k,i+1} \\ D_{k,i} & \text{otherwise} \end{cases}$$

Now, let

$$\mathrm{T}_{k+1} = \mathrm{T}_{k,n_k{}^{l_k}}$$

and

$$D_{k+1} = D_{k,n_k{}^{l_k}}$$

Finally, put

$$\mathrm{T}_\infty = \bigcup_k \mathrm{T}_k$$

and

$$D_\infty = \bigcup_k D_k$$

[16] The kind of recursion used here is allowed since we do not violate the condition of stratification explained on page 145.

To show that T_∞ is consistent, we prove the stronger claim that T_∞ is consistent with $D^{\neq}_\infty$. This relative consistency is shown by induction over k in

$$\mathrm{T}_k \text{ is consistent with } D^{\neq}_k$$

When $k = 0$, $\mathrm{T}_k = \mathrm{T}$ and $D^{\neq}_k = \emptyset$, so the claim holds. Assume that the claim holds for some k. We show that it holds also for $k + 1$, this time, by induction over i in

$$\mathrm{T}_{k,i} \text{ is consistent with } D^{\neq}_{k,i}$$

Clearly, this holds for $i = 0$ (by assumption). Assume it holds for some i and consider $\theta_{k,i}$. In order to show that $\mathrm{T}_{k,i+1}$ is consistent with $D^{\neq}_{k,i+1}$ we need to consider a number of cases:

1. $\theta_{k,i}$ is of the form $\exists x\varphi$.
   a) There is a $d \in D_{k,i}$ such that $\mathrm{T}'_{k,i} \cup \{\varphi(d)\}$ is consistent with $D^{\neq}_{k,i}$. Then $\mathrm{T}_{k,i+1} = \mathrm{T}'_{k,i} \cup \{\varphi(d)\}$ and $D_{k,i+1} = D_{k,i}$, which gives the result.
   b) There is no $d \in D_{k,i}$ such that $\mathrm{T}'_{k,i} \cup \{\varphi(d)\}$ is consistent with $D^{\neq}_{k,i}$. Then we have two possibilities:
      i. There is a $d' \notin D_{k,i}$ such that $\mathrm{T}'_{k,i} \cup \{\varphi(d')\}$ is consistent with $(D_{k,i} \cup \{d'\})^{\neq}$. Then $\mathrm{T}_{k,i+1} = \mathrm{T}'_{k,i} \cup \{\varphi(d')\}$ and $D_{k,i+1} = D_{k,i} \cup \{d'\}$, which gives the relative consistency.
      ii. There is no d' such that $\mathrm{T}'_{k,i} \cup \{\varphi(d')\}$ is consistent with $(D_{k,i} \cup \{d'\})^{\neq}$. Then $\mathrm{T}_{k,i+1} = \mathrm{T}'_{k,i}$ and $D_{k,i+1} = D_{k,i}$. Thus, the relative consistency follows from the construction of $\mathrm{T}'_{k,i}$.
2. $\theta_{k,i}$ is not of the form $\exists x\varphi(x)$. Then, again, $\mathrm{T}_{k,i+1} = \mathrm{T}'_{k,i}$ and $D_{k,i+1} = D_{k,i}$ so the relative consistency follows from the construction of $\mathrm{T}'_{k,i}$.

Thus, if T_k is consistent with D_k, then $\mathrm{T}_{k,i}$ is consistent with $D^{\neq}_{k,i}$ for all i, and it follows that T_{k+1} is consistent with $D^{\neq}_{k+1}$. Hence, T_∞ is consistent with $D^{\neq}_\infty$, so T_∞ is consistent.

T_∞ is maximal if, for each $\mathscr{L}_{T_\infty}$-formula θ, either θ or $\neg\theta$ belongs to T_∞. Assume thus that $\theta \notin T_\infty$, meaning that, for all k, $\theta \notin T_k$. Now, for some k and m, θ is $\theta_{k,m}$. Since $T'_{k,m} \subseteq T_{k,m+1}$ and $\theta_{k,m} \notin T_{k,m+1}$, it follows that $\theta_{k,m} \notin T'_{k,m}$. But then, by the construction of $T'_{k,m}$, $\neg\theta_{k,m} \in T'_{k,m}$, which, again by the construction, gives that $\neg\theta \in T_\infty$.

The witness completeness of T_∞ follows directly from the construction. Thus, T_∞ is maximal, consistent and witness complete.

Next we want to construct a model $\mathcal{M}$ such that, for each $\mathscr{L}_{T_\infty}$-sentence θ, $T_\infty \vdash \theta$ iff $\mathcal{M} \models \theta$. If we put $\mathcal{M} = \langle D_\infty, I\rangle$ the problem turns into a problem of defining I. Define:

1. for all $d \in D_\infty$, $\langle d, d\rangle \in I$
2. for all $c \in \mathscr{L}_{\text{CONS}}$, let $\langle c, d\rangle \in I$ for the unique $d \in D_\infty$ s.t. $T_\infty \vdash c = d$
3. For all $P^n \in \mathscr{L}_{\text{PRED}}$ let, for all $d_{i_1}, \ldots, d_{i_n} \in D_\infty$, $\langle P^n, \langle d_{i_1}, \ldots, d_{i_n}\rangle\rangle \in I$ if, and only if $T_\infty \vdash P^n(d_{i_1}, \ldots, d_{i_n})$
4. For all $f^n \in \mathscr{L}_{\text{FUNC}}$ let, for all $d_{i_1}, \ldots, d_{i_n}, d_{i_{n+1}} \in D_\infty$, $\langle f^n, \langle d_{i_1}, \ldots, d_{i_n}, d_{i_{n+1}}\rangle\rangle \in I$ if, and only if $T_\infty \vdash f^n(d_{i_1}, \ldots, d_{i_n}) = d_{i_{n+1}}$.

$T_\infty \vdash \theta$ iff $\mathcal{M} \models \theta$ is now proved by induction on the complexity of θ, which shows that $\mathcal{M}$ is a model for T_∞.

We restrict I to $\mathscr{L}_T$ to get $\langle D_\infty, I \restriction \mathscr{L}_T\rangle$, which is a model for T. □

Corollary 6.2.15. $T \models \varphi \Rightarrow T \vdash \varphi$

Corollary 6.2.16. Let T be a theory in a denumerable language. If T has model, then it has a denumerable model.

6.3 Concluding remarks

The construction of a model-theoretic semantics in NFU_p for absolute quantification is a comforting result. It gives us the means to model the semantics for any first-order language in such a way that the quantifiers are interpreted to range over absolutely everything. Moreover, since each semantic concept is defined by a stratified formula, there is a model for the semantic theory itself, formulated as a first-order theory. This takes us as close as we can get to a first-order formulation of a strictly adequate model-theoretic semantics for absolute quantification.

We cannot, in $\mathrm{NFU_p}$, hope for a model $\langle V, I\rangle$ for $\mathscr{L}_{\mathrm{NFU_p}}$ such that $\langle V, I\rangle \models \mathrm{NFU_p}$ since the complement of the intersection of the image of I under $\{\in\}$ and the identity set, i.e. the set of all pairs $\langle x, x\rangle$, would be the set of all pairs $\langle x, x\rangle$ such that $x \notin x$, which we know doesn't exist. But this is not surprising. To construct a model-theoretic semantics we need to pick a metatheory T to work in and by Gödel's results we cannot, in T, prove the existence of a model of T since that would constitute a proof of the consistency of T.[17]

Just as Williamson suggests that we should learn to use higher-order languages as our home language we would suggest that one should use $\mathscr{L}_{\mathrm{NFU_p}}$ as our home language.[18] And just as Williamson cannot provide a semantics for the hierarchy of languages he suggests, we cannot provide a model for $\mathscr{L}_{\mathrm{NFU_p}}$ in $\mathrm{NFU_p}$. But whereas Williamson needed to go beyond natural languages for semantic theorising, $\mathscr{L}_{\mathrm{NFU_p}}$ is a first-order language that can easily be couched in natural language.

The lack of an extension of $\in$ in $\mathrm{NFU_p}$ is reminiscent of Linnebo's lack of a property for η. We too must yield to the fact that we may not provide the intended model for our adopted metalanguage within our constructed semantics. This shows that the semantics fails to be syntactically strictly adequate, but it is, which is a remarkable fact, semantically strictly adequate with regard to sets as extensions of predicates.[19] In particular, it resists the argument from cardinality. The reason is that Cantor's theorem does not hold for $|V|$ in $\mathrm{NFU_p}$, i.e., we do not have $|V| < |\wp V|$. Moreover, because of fact 6.2.13, the constructed semantics is not vulnerable to WA.

Finally, the completeness result shows that logical consequence behaves in the right way for first-order languages in our meta-theory.

[17] That is, assuming that T is, in fact, consistent.

[18] We discussed Williamson's suggestion in Chapter 5.

[19] As the reader recalls, we said on page 90 that semantic adequacy entails syntactic adequacy if all contentful predicates have semantic values. In this case the antecedent fails for the predicate for the membership relation and thus, so does the entailment.

7 Concluding Remarks and Further Questions

Let us recapitulate. In Chapter 2 we discussed the theory of quantification in *Begriffsschrift* and the notion of indefinite extensibility in *Grundgesetze*. Indefinite extensibility has been taken to warrant the impossibility of absolute quantification under the assumption that the quantifiers are always restricted to a domain of quantification. But the commitment to a domain of quantification as an entity over and above the objects quantified over is, it may be argued, untenable. We said that the adoption of a model-theoretic semantics, although it takes sets as domains of quantification, imposes no new ontological commitments in addition to the objects quantified over in the object languages. The key observation is that a model-theoretic semantics uses sets to *represent* entities and structures without any claims of *being* those entities and structures.

In Chapter 3 we discussed at some length the possibility of adopting a set-theoretic version of Tarski's domain free truth definition in CTFL for absolute quantification. However, the requirement that our semantics must allow for a definition of logical consequence made it impossible to carry out the definition in a set theory with a well-founded membership relation.

In Chapter 4 WA was discussed in detail. The main conclusion of that chapter was that WA shows a particular definition to be faulty, rather than constituting a reductio of absolute quantification. Moreover, no independent reason for the correctness of that definition has been given.

In Chapter 5 we critically reviewed three theories that in different ways meet Willaimson's challenge of providing an adequate semantics for absolute quantification. Two of these theories adopt a type-theoretic framework which is mainly motivated in three ways. First, WA is understood to say that first-order quantification over interpretations (of first-order languages) is paradoxical. Second, a variant of Cantor's theorem is taken to show that there must be more interpretations of predicates than individuals in order for a semantics to be strictly adequate which, in turn, is taken to show that we cannot use first-order quantification over individuals to generalise over interpretations. Third, for a given language in the hierarchy it is possible to show that there is a language with greater expressive power higher up in the hierarchy that allows for a strictly adequate semantics for the first language.

The third theory we discussed, Linnebo's theory on sets and properties, resists the move to type-theoretic languages by adopting a restrictive theory of properties that makes a premise of WA false.

We saw that the type-theorists face some difficulties. Since natural language lacks higher-order quantifiers, concepts, and plural terms, it turns out to be hard for the type-theorists to give a stringent presentation of their semantics. Moreover, as Linnebo reminded us, certain concepts naturally cut across types and are thus not available in a type-theoretic framework. This holds for important semantic concepts, but also, as we saw, for the concept of strict adequacy. We analysed strict adequacy into a semantic and a substitutional version. The substitutional version was seen to give rise to a Grelling type paradox which constitutes a dilemma for the type-theorist: either he blocks the Grelling type paradox in a way that will block WA as well, or his semantic theory will be inadequate in the substitutional sense. Linnebo's approach, which is not type-theoretic, is not affected by these arguments.

Very much in sympathy with Linnebo's approach, but in spite of his reservations about NF style set theory, we argued in Chapter 6 that it is possible use a non-standard set theory to develop a semantics for absolute quantification. We saw how to construct a respectable first-order model-theoretic semantics that allows for domains of quantification containing absolutely everything. The semantics is strictly adequate in the semantic sense. That is, for any possible semantic value of a predicate, there is a model interpreting the predicate accordingly. It is not, however, strictly adequate in the syntactic sense. In particular, the membership relation cannot be given its intended extension. Thus, what we may claim to have shown is that, given the membership relation and, for convenience, the pairing operator, as primitives, one can construct a strictly adequate model-theoretic semantics for first-order languages with absolute quantification.

In particular, the concepts of the semantics may themselves be given their intended interpretations in the semantics constructed.

7.1 Further questions

It is important to note in this context that stratified comprehension is not by itself restrictive as to the existence of classes. Any class that has a stratified definition exists and any alleged class lacking a stratified definition may exist *unless we have an argument why it shouldn't exist.* As long as the resulting theory is consistent we are free to add classes that lack stratified definitions. Indeed, Feferman (1974, 2011) uses an extension S^* of $\mathrm{NFU_p}$ that has ZFC as a sub-theory. The idea is to add to $\mathscr{L}_{\mathrm{NFU_p}}$ a constant for the ZFC-universe and add variables ranging over this universe that may have different types at different occurrences in the same stratified formula. It is an interesting question if Linnebo's theory can be interpreted by our semantics as developed in S^*, or vice versa.

A second issue of interest concerns the consequence relation. A worry that has been expressed concerning the possibility of developing a model-theoretic semantics in $\mathrm{NFU_p}$ is that, since not every subclass of V is a set, the consequence relation would be inadequate. Corollary 6.2.15 proves that this worry is uncalled-for in the first-order case. However, for other cases, the relationship between the standard consequence relation and the one defined in $\mathrm{NFU_p}$ remains to be investigated. Such an investigation is not completely straightforward. For instance, to define INT to include models interpreting, e.g., second-level predicates, i.e., predicates taking second-order variables as arguments, by letting I hold between such predicates and subsets of the domain would make the definition unstratified. Thus, in order to compare the consequence relations in $\mathrm{NFU_p}$ and ZFC for higher-order languages, some semantics in $\mathrm{NFU_p}$ for higher-order languages need to be constructed. That work remains to be done.

Finally, once we have the model-theoretic semantics for absolute first-order quantification in place we may start asking questions about what it would mean to have the domain of absolutely everything available as a domain of inquiry. Say that the extension of a formula φ with one free variable in a model $\mathcal{M}$, $\varphi^{\mathcal{M}}$, is the set of all objects in the domain of $\mathcal{M}$ for which φ is true in $\mathcal{M}$. Then we may define a quantifier corresponding to 'absolutely everything', Q_V, by

$$\mathcal{M} \models Q_V x \varphi(x) \text{ if and only if } \varphi^{\mathcal{M}} = V$$

It is then natural to proceed to investigate what it would mean to add Q_V to first-order logic.

References

Baker, G. (2001). 'Function' in Frege's *Begriffsschrift*: Dissolving the problem. *Brittish Journal for the History of Philosophy*, 9(3):525–544.

Baker, G. P. and Hacker, P. M. S. (1984). *Frege : Logical Excavations*. Oxford U.P., New York.

Baker, G. P. and Hacker, P. M. S. (2003). Functions in *Begriffsschrift*. *Synthese*, 135(3):273–297.

Beaney, M., editor (1997). *The Frege Reader*. Blackwell Publishing Ltd.

Bennet, C. and Filin Karlsson, M. (2008). Williamson's barber. *Analysis*, 68(4).

Boolos, G. (1971). The iterative conception of set. *The Journal of Philosophy*, 68(8):215–231. Reprinted in Boolos (1998), 13–29.

Boolos, G. (1984). To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy*, 81(8):430–449. Reprinted in Boolos (1998), 54–72.

Boolos, G. (1985). Nominalist platonism. *The Philosophical Review*, 94(3):327–344. Reprinted in Boolos (1998), 73–87.

Boolos, G. (1998). *Logic, Logic and Logic*. Harvard University Press, paperback edition.

Cartwright, R. (1994). Speaking of everything. *Noûs*, 28(1):1–20.

Church, A. (1974). Set theory with a universal set. In *Proceedings of the Tarski symposium*, volume 25, pages 297–308. Proc. Symposia Pure Math. XXV, AMS, Providence RI.

David, M. (2008). Tarski's convention T and the concept of truth. In Patterson, D., editor, *New essays on Tarski and philosophy*. Oxford University Press, Oxford.

Dummett, M. (1984). An unsuccessful dig. *The Philosophical Quarterly*, 34(136):377–401.

Dummett, M. (1991). *Frege: Philosophy of Mahematics*. Harvard University Press.

Dunn, J. M. and Belnap, Jr, N. D. (1968). The substitution interpretation of the quantifiers. *Noûs*, 2(2):177–185.

Feferman, S. (1974). Some formal systems for the unlimited theory of structures and categories. Unpublished MS, available at `http://math.stanford.edu/~feferman/papers.html`.

Feferman, S. (2008). Tarski's conceptual analysis of semantical notions. In Patterson, D., editor, *New essays on Tarski and philosophy*. Oxford University Press, Oxford.

Feferman, S. (2011). Enriched stratified systems for the foundations of category theory. In Sommaruga, G., editor, *Foundational Theories of Classical and Constructive Mathematics*, pages 127–143. Springer Netherlands, Dordrecht.

Forster, T. (1995). *Set Theory with a Universal Set: Exploring an Untyped Universe*. Clarendon press (Oxford), second edition.

Forster, T. (2008). The iterative conception of set. *Review of Symbolic Logic*, 1(1):97–110.

Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. L. Nebert. Reprinted and translated to English by S. B. Mengleberg in van Heijenoort (1967), pp. 5–82.

Frege, G. (1891). Funktion und Begriff. *Vortrag, gehalten in der Sitzung vom*, 9. Translated to English by P. Geach and M. Beaney as Function and concept in Beaney (1997), pages 130–148.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50. Translated to English by M. Black as On Sinn and Bedeutung in Beaney (1997), pages 151–171.

Frege, G. (1893,1903). *Grundgesetze der Arithmetik*, volume 1–2. Hermann Pohle. Partially translated to English by M. Furth as *The Basic Laws of Arithmetic*, Berkeley: University of California Press, 1964.

Glanzberg, M. (2004). Quantification and realism. *Philosophy and Phenomenological Research*, LXIX(3):541–572.

Heck, R. (2003). Frege on identity and identity-statements: A reply to Thau and Caplan. *Canadian Journal of Philosophy*, 33(1):83–102.

Hodges, W. (1985/6). Truth in a structure. *Proceedings of the Aristotelian Society*, New series 86:135–152.

Hodges, W. (2004). What languages have Tarski truth definitions? *Annals of Pure and Applied Logic*, 126:93–113.

Holmes, R. (1998). *Elementary set theory with a universal set*. Cahiers du Centre de logique, Volume 10, Academia, Louvaine-la-Neuve. `http://math.boisestate.edu/%7Eholmes/holmes/head.pdf`

Jensen, R. B. (1968-69). On the consistency of a slight (?) modification of Quine's *New Foundations*. *Synthese*, 19:250–263.

Kreisel, G. (1967). Informal rigour and completeness proofs. *Studies in Logic and the Foundations of Mathematics*, 47:138–186.

Lavine, S. (2006). Something about everything: Universal quantification in the universal sense of universal quantification. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 98–148. Oxford University Press.

Linnebo, Ø. (2003). Plural quantification exposed. *Noûs*, 37(1):71–92.

Linnebo, Ø. (2006). Sets, properties and unrestricted quantification. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 149–178. Oxford University Press.

Linnebo, Ø. (2012). Plural quantification. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2012 edition.

Linnebo, Ø. and Nicolas, D. (2008). Superplurals in English. *Analysis*, 68(3):186–197.

Maddy, P. (1997). *Naturalism in mathematics*. Clarendon Press, Oxford.

Marcus, R. B. (1972). Quantification and ontology. *Noûs*, 6(3):240–250.

May, R. (2001). Frege on identity statements. In Cecchetto, C., Chierchia, G., and Guasti, M. T., editors, *Semantic Interfaces: Reference, Anaphora, and Aspect*, pages 1–51. CSLI Publications.

McGee, V. (2000). 'Everything'. In Sher, G. and Tieszen, R., editors, *Between logic and intuition*, pages 54–78. Cambridge University Press.

Parsons, C. (1982). Objects and logic. *The Monist*, 65(4):491–516.

Parsons, C. (1990). The structuralist view of mathematical objects. *Synthese*, 84(2):303–346.

Parsons, C. (2006). The problem of absolute universality. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 203–219. Oxford University Press.

Patterson, D. (2008). *New Essays on Tarski and Philosophy*. Oxford University Press.

Peters, S. and Westerståhl, D. (2006). *Quantifiers in Language and Logic*. Oxford University Press.

Quine, W. v. O. (1937). New foundations for mathematical logic. *The American Mathematical Monthly*, 44(2):70–80. in Quine (1961), pages 80–101.

Quine, W. v. O. (1941). Whitehead and the rise of modern logic. In Schilpp, P. A., editor, *The Philosophy of Alfred North Whitehead*, pages 125–164. La Salle: Open Court.

Quine, W. v. O. (1945). On ordered pairs. *Journal of Symbolic Logic*, 10(3):95–96.

Quine, W. v. O. (1948). On what there is. *The Review of Metaphysics*, 2(1):21–38. Reprinted in Quine (1961), pages 1–19.

Quine, W. v. O. (1961). *From a logical point of view: 9 logico-philosophical essays*, volume 9. Harvard University Press, second revised edition.

Quine, W. v. O. (1968). Ontological relativity. *The Journal of Philosophy*, 65(7):185–212.

Quine, W. v. O. (1970). *Philosophy of Logic*. Prentice-Hall Foundations of Philosophy Series. Prentice-Hall, Inc.

Quine, W. v. O. (1995). The inception of "new foundations". In *Selected logical papers*, pages 286–289. Enlarged edition.

Rayo, A. (2002). Word and objects. *Noûs*, 36:436–464.

Rayo, A. (2006). Beyond plurals. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 220–254. Oxford University Press.

Rayo, A. (2012). Absolute generality reconsidered. *Oxford Studies in Metaphysics*, 7:93–125.

Rayo, A. (2017). The world is the totality of facts, not of things. *Philosophical Issues*, 27(1):250–278.

Rayo, A. and Uzquiano, G. (1999). Toward a theory of second-order consequence. *Notre Dame Journal of Formal Logic*, 40:315–325.

Rayo, A. and Uzquiano, G., editors (2006). *Absolute Generality*. Clarendon Press Oxford.

Rayo, A. and Williamson, T. (2003). A completeness theorem for unrestricted first-order languages. In Bell, J., editor, *Liars and Heaps: New Essays on Paradox*, chapter 15, pages 331–356. Clarendon Press Oxford.

Rayo, A. and Yablo, S. (2001). Nominalism through de-nominalization. *Noûs*, 35(1):74–92.

Resnik, M. D. (1988). Second-order logic still wild. *The Journal of Philosophy*, 85(2):75–87.

Rosser, J. B. (1952). The axiom of infinity in Quine's New Foundations. *The Journal of Symbolic Logic*, 17(4):238–242.

Russell, B. (1907). On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society*, s2-4(1):29–53.

Russell, B. (1908). Mathematical logic as based on the theory of types. *American journal of mathematics*, 30(3):222–262. Reprinted in van Heijenoort (1967), pages 152–182.

Russell, B. (1914). *Our Knowledge of the External World: As a field for scientific method in philosophy*. George Allen & Unwin LTD. Revised and reset 1926, Reprinted 1949.

Russell, B. and Whitehead, A. N. (1910). *Principia Mathematica*, volume 1. Cambridge University Press.

Russinoff, S. (1987). On the brink of a paradox? *Notre Dame Journal of Formal Logic*, 28(1):115–131.

Shapiro, S. and Wright, C. (2006). All things indefinitely extensible. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 255–304. Oxford University Press.

Simons, P. (1997). Higher-order quantification and ontological commitment. *Dialectica*, 51:255–271.

Simons, P. (2001). Categories, construction and congruence: Husserl's tactics of meaning. In Gaskin, R., editor, *Grammar in early twentieth-century philosophy*, pages 152–278. Routledge.

Smith, P. (2011). Squeezing arguments. *Analysis*, 71(1):22–30.

Specker, E. (1953). The axiom of choice in Quine's new foundations for mathematical logic. *Proceedings of the National Academy of Sciences*, 39(9):972–5.

Studd, J. (2017). Generality, extensibility, and paradox. In *Proceedings of the Aristotelian Society*, volume 117, pages 81–101. Oxford University Press.

Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1. Translated to English by J. H. Woodger as 'The concept of truth in formalized languages' in Tarski (1956).

Tarski, A. (1936a). Grundlegung der wissenschaftlichen Semantik. *Actes du Congrès International de Philosophie Scientifique*, 3. Translated to English by J. H. Woodger as 'The establishment of scientific semantics' in Tarski (1956).

Tarski, A. (1936b). Über den Begriff der logishen Folgerung. *Actes du Congrès International de Philosophie Scientifique*, 7. Translated to English by J. H. Woodger as 'On the concept of logical consequence' in Tarski (1956).

Tarski, A. (1944). The semantic conception of truth: and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3):341–376.

Tarski, A. (1956). *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Hackett Pub., second edition.

Tarski, A. and Vaught, R. (1957). Arithmetical extensions of relational systems. *Compositio Mathematica*, 13:81–102.

van Heijenoort, J., editor (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press Cambridge, MA. Third printing (1977).

Williamson, T. (2003). Everything. In Hawthorne, J. and Zimmerman, D., editors, *Philosophical Perspectives 17:1 Language and Philosophical Liguistics*, pages 415–65. Blackwell.

Zermelo, E. (1908). Untersuchungen über die Grundlagen der Mengenlehre. I. *Mathematische Annalen*, 65(2):261–281. Translated by S. B. Bauer-Mengelberg as 'Investigations in the foundations of set theory' in van Heijenoort (1967), pages 201–215.

Sammanfattning

Sällan har vi för avsikt att prata om absolut allt som finns. Tvärtom, i våra vardagskonversationer är kvantifikation nästan alltid begränsad på ett eller annat vis. Begränsningarna kan vara implicita, till exempel via en kontextuellt given bakgrundsdomän, eller explicita via syntaktiska mekanismer. Men det verkar samtidigt finnas tillfällen då vi faktiskt strävar efter att kvantifiera över absolut allting. Bara en illvillig lyssnare skulle förstå kvantifikatorn som begränsad om en metafysiker påstod att ”Allting tillhör någon ontologisk kategori”. På samma sätt gäller att en mängdteoretiker som förklarar att ”Ingenting är element i den tomma mängden” inte menar att kvantifikatorn bara gäller för en begränsad domän utanför vilken det kan finnas potentiella element som skulle kunna göra den tomma mängden icke-tom trots allt. Ett tredje exempel är Aristoteles identitetslag – för allting A gäller att A är A – som är tandlös om kvantifikatorn vore begränsad.

Men även om kvantifikation över absolut allting ter sig oproblematisk i vardagsspråket är sådan kvantifikation likväl behäftad med intrikata och svåra filosofiska problem. De mest utmanande härrör från de matematisk-logiska paradoxerna. Så har Cantors paradox om det största kardinaltalet, Burali-Fortis paradox om det största ordinaltalet och Russells klassparadox alla använts i försök att visa att själva idén med absolut kvantifikation är inkoherent. Argument av detta slag antar vanligen att kvantifikation alltid förutsätter en domän bestående av de ting över vilka vi kvantifierar. Resonemangen i paradoxerna används för att visa att varje sådan domän kan expanderas till en mer omfattande domän och på så vis kan det inte finnas någon största domän, än mindre en domän av absolut allting. Dummett (1991) har kallat de begrepp med vilka vi bestämmer sådana domäner för indefinit expanderbara och Russell (1907) har kallat de klasser, eller extensioner, som hör till de begreppen, för självreproduktiva. Enligt argument av detta slag finns det alltså ingen universell domän och följaktligen heller inget sådant som absolut kvantifikation.

En möjlig reaktion på argument som utgår från indefinit expanderbarhet är att utmana intuitionerna om kvantifikatorerna i exemplen ovan och acceptera slutsatsen att kvantifikation alltid är begränsad till något som är mindre omfångsrikt än totaliteten av absolut allting. En sådan position kallar vi *generalitetsrelativism*. Motsatt position, att kvantifikation kan vara sant universell, benämns *generalitetsabsolutism*. Vi använder 'relativist' respektive 'absolutist' för respektive positions försvarare.

Trots att generalitetsrelativismen kan tyckas vara en naturlig position i ljuset av indefinit expanderbarhet så är den samtidigt djupt problematisk. Timothy Williamson visar i sin tankväckande och inflytelserika *Everything* (2003) att, inte bara är relativisten oförmögen att artikulera sin egen position på ett koherent sätt, han är också oförmögen att ge adekvata teorier om generaliseringar över sorter och, vilket är viktigare, mening och sanning. Williamson visar vidare att, givet vissa naturliga antaganden om kontexter och språk, kan relativisten inte formulera sanningsvillkoren för en kontextkänslig universellt kvantifierad utsaga i ett kontextkänsligt metaspråk.

Det relativisten önskar säga, enligt Williamson, är

(∗) för varje kontext C, och varje sats på formen $\forall x\varphi$, gäller att $\forall x\varphi$ är sann i C om och endast om varje element i C:s domän satisfierar φ i C.

Eftersom kvantifikation i metaspråket är kontextkänsligt gäller att den kontext i vilken (∗) yttras, låt oss kalla den CT, har en domän. Williamson påpekar att för en kontext C som instantierar (∗) erhåller vi villkoret att $\forall x\varphi$ är sann i C om och endast om varje element i snittet av C:s domän och CT:s domän satisfierar φ i C. Alltså, för att (∗) ska ge rätt sanningsvillkor för varje kontext C måste CT:s domän innehålla varje element i unionen av domänerna för de C i vilken $\forall x\varphi$ kan yttras. Men ett krav på en sådan kontext ligger farligt nära ett krav på en kontext vars domän inbegriper absolut allting. Alternativet att det finns något objekt som inte tillhör någon av domänerna i de möjliga kontexterna för $\forall x\varphi$ skulle innebära att CT:s domän inte behöver inbegripa absolut allting. Men att det skulle finnas ett sådant objekt, om $\forall x\varphi$ är en sats i naturligt språk, ter sig osannolikt och det är hursomhelst något som relativisten inte kan uttrycka.

Detta och liknande argument får Williamson att hävda att generalitetsrelativism leder till en slags metalingvistisk pessimism i meningen att den undergräver möjligheten till en reflexiv förståelse av vårt eget tänkande och språk, även från ett metaspråkligt perspektiv.[1] Uppgiften för absolutisten blir att visa att givet hans förståelse av kvantifikation, så är en koherent metalingvistisk reflektion faktisk möjlig.

Williamson hävdar att vanlig modellteoretisk semantik kan kritiseras på samma grunder: i metaspråket för första ordningens modellteori säger vi att varje modell har en mängd som domän; men varje objekt tillhör någon mängd (så som sin egen enhetsmängd) och därför också någon domän till en eller annan modell. Det följer att ingen modell har varje modell i sin domän. Alltså saknar varje formalisering av metateorin i ett första ordningens språk en standardmodell.[2] Enligt Williamson är modellteoretisk semantik i sin standardformulering alltså inte bara inadekvat för absolut kvantifikation på grund av avsaknaden av en universell mängd, utan också eftersom det inte finns en standardmodell för teorin själv.

I avhandlingen argumenterar jag för att vi inte behöver ge upp idén om en första ordningens modellteoretisk semantik för absolut kvantifikation. Absolutisten kan mycket väl formulera en modellteoretisk semantik som inte drabbas av Williamsons kritik. En sådan formulering kommer innehålla ”modell”, ”satisfierar” och ”evaluering” bland sina predikat. En modell, $\mathcal{M}^{\Pi}$, för ett sådant språk kommer, likt varje modell för ett första ordningens språk, innehålla en domän, M^{Π}, för kvantifikation och en funktion I^{Π} som tolkar predikaten i språket. I kapitel 6 definierar jag $\mathcal{M}^{\Pi}$ i mängdteorin $\mathrm{NFU_p}$ som är Quines NF med urelement och en primitiv paroperator. De modeller i M^{Π} som har den universella mängden som sin domän är modeller för absolut kvantifikation. Semantiken som definierats visas vara fullständig med avseende på vedertagna klassiska bevissystem. Sålunda är begreppet *första ordningens konsekvens* i den nya semantiken extensionellt ekvivalent med begreppet *härledbar*, vilket i sin tur implicerar att det är extensionellt ekvivalent med första ordningens konsekvens i vanlig modellteoretisk semantik. Alltså avviker inte den i $\mathrm{NFU_p}$ definierade modellteoretiska semantiken väsentligen från vanlig modellteoretisk semantik.

[1] Williamson (2003, s. 452).
[2] Williamson (2003, s. 446).

Men att vi *kan* definiera en semantik i NFU_p implicerar inte att vi *bör* göra det och det är klart att vår nya semantik behöver motiveras ytterligare. Kapitel 2–5 är ämnade att ge en slags motivering, delvis genom att kommentera diskussioner och motargument som återfinns i litteraturen. Nedan följer en kort summering av de viktigaste diskussionerna i respektive kapitel.

En naturlig utgångspunkt för ett arbete om semantiken för absolut kvantifikation är Gottlob Freges arbeten. En anledning är att han införde ett logiskt system i sin *Grundgesetze der Arithmetik* (1893,1903) i vilket Russells paradox kan härledas. Sålunda försåg han oss (oavsiktligt) med ett av de mest inflytelserika argumenten mot absolut kvantifikation. Ytterligare ett skäl är att han, enligt gängse läsningar, begagnade, eller avsåg att begagna, absolut kvantifikation genom att låta sina första ordningens kvantifikatorer kvantifiera över absolut alla objekt.

Frege var klar över syntaxen och i viss utsträckning även semantiken för kvantifikation redan i sin första bok om logik, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* (1879). I första delen av kapitel 2 utmanar jag den allmänt vedertagna uppfattningen att Frege ämnade kvantifiera över absolut allting i det verket. Närmare bestämt argumenterar jag för att kvantifikatorerna däri bäst förstås i termer av substitution. Det vill säga, istället för att $\forall x \varphi$ är sann om och endast om φ är sann för alla *värden* på x, så hävdar jag att kvantifikatorerna i *Begriffsschrift* gör den sann om och endast om φ är sann för alla legitima substitutionsinstanser av x. Då vi endast kvantifierar över namngivna objekt – om vi överhuvud taget kvantifierar *över* något – vid substitutionskvantifikation, så räcker det att anta existensen av ett objekt som inte denoteras av något uttryck i *Begriffsschrift*, för att vederlägga absolut kvantifikation däri.

I andra delen av kapitel 2 diskuterar jag Dummetts inflytelserika argument, som stammar ur hans analys av *Grundgesetze*, att absolut kvantifikation är inkoherent då det finns indefinit expanderbara begrepp. Richard Cartwrights *Speaking of Everything* (1994) ger ett intressant svar på detta argument. Cartwright gör gällande att det är missvisande att anta, som Dummett gör, att det måste finnas en definit kollektion av de objekt vi kvantifierar över utöver objekten själva. Cartwrights svar är intressant i sig, men det ger oss också skäl att diskutera relationen mellan modellteoretisk semantik och ontologiska åtaganden i objektspråket. Ett potentiellt problem som tycks följa ur att vi kvantifierar över, och därmed antar existensen av, domäner i metaspråket är att ett objektspråk med absolut kvantifikation verkar ärva metaspråkets ontologiska åtagande till domäner. Därmed tycks det som att vi trots allt, via en omväg i metaspråket, förutsätter just en definit kollektion av de objekt vi kvantifierar över när vi i objektspråket kvantifierar över absolut allting. Vi avslutar kapitel 2 med en diskussion som reder ut denna fråga.

För att konstruera en modellteoretisk semantik behöver vi definiera relationen, $\mathcal{M} \models \varphi$, som råder mellan en sats φ och en modell $\mathcal{M}$ när φ är sann i $\mathcal{M}$. Alfred Tarski var först med att ge en formell definition av denna relation, men han gav den inte, som det ibland påstås, i *The Concept of Truth in Formalized Languages* (1935).[3] Däri definierar Tarski sanning för formaliserade språk, alltså språk som liknar formella språk i det att de har en precis syntax men skiljer sig från dem i det att de är meningsfulla utan att stå i en relation till en tolkning eller modell. En intressant aspekt av Tarskis definition är att den inte innehåller någon explicit begränsning på kvantifikatorerna. Frågan om det är möjligt att modifiera hans metod att definiera sanning till att passa språk med absolut kvantifikation är därför naturlig. Jag diskuterar den frågan i första delen av kapitel 3 och visar att Tarskis användning av Husserls semantiska kategorier i metaspråket omöjliggör en sådan modifikation. Eftersom variablerna (i metaspråket) för evalueringar med nödvändighet tillhör en annan semantiska kategori än variablerna i objektspråket, verkar det i princip omöjligt att kvantifiera över evalueringar i objektspråket. Alltså, från det metaspråkliga perspektivet finns det något som vi inte kvantifierar över i objektspråket, och följaktligen kvantifierar vi inte över absolut allting.

I den andra delen av kapitel 3 föreslås två sätt att komma runt problemet som Husserls kategorier ger upphov till i samband med absolut kvantifikation. Det första alternativet gör gällande att Husserls semantiska kategorier inte nödvändigtvis behöver innebära nya ontologiska åtaganden i metasprå-

[3] Den första tryckta definitionen av begreppet *sann i en modell* tycks förekomma i Tarski och Vaught (1957). Se Hodges (1985/6) för en diskussion.

ket. Detta alternativ ligger nära typhierarkiska semantiska teorier som söker undvika en ontologi utöver den som är given i den lägsta typen. Jag diskuterar två sådana teorier i kapitel 5. Den andra lösningen består i att ersätta Husserls kategorier med mängdteoretiska konstruktioner.

Intresset för modellteoretisk semantik förklaras till stor del av dess förmåga att ge adekvata definitioner av logiska relationer mellan satser. Till exempel är den modellteoretiska definitionen av logisk konsekvens av fundamental betydelse. I tredje delen av kapitel 3 visar jag att semantiken som följer av att ersätta Husserls kategorier med ZF, även om den ger en korrekt sanningsdefinition, inte ger en definition av logisk konsekvens. En intressant orsak till detta är regularitetsaxiomet, som gör $\in$ välgrundad. I kapitel 6 använder vi istället $\mathrm{NFU_p}$ vars $\in$-relation inte är välgrundad.

Dummetts argument, att indefinit expanderbara begrepp visar på inkoherensen hos absolut kvantifikation, är formulerat i en mängdteoretisk anda och vilar på antaganden om mängder och klasser. Ett liknande argument, som däremot inte gör några antaganden om mängder eller klasser, ges av Williamson (2003). Argumentet presenteras som ett *reductio* av antagandet om absolut kvantifikation. Utöver antagandet om absolut kvantifikation används ytterligare två premisser. Den första gör gällande att, givet ett 1-ställigt predikat i objektspråket, så finns, för varje möjligt semantiskt värde på predikatet, en tolkningsfunktion som tilldelar predikatet det semantiska värdet ifråga; särskilt gäller detta för de semantiska värdena som predikaten i (det tolkade) metaspråket antas ha. En semantik med denna egenskap sägs vara *strikt adekvat*. Den andra premissen är att en viss definition av ett metaspråkligt predikat är legitim. Även om argumentet inte gör några antaganden om mängder eller klasser har det analyserats i termer av indefinit expanderbarhet av till exempel Glanzberg (2004) och Parsons (2006).

Williamsons argument har en central roll i den samtida diskussionen om absolut kvantifikation och hela kapitel 4 ägnas åt det. Efter att ha presenterat argumentet diskuterar jag Glanzbergs och Parsons olika analyser. Jag ger också en alternativ analys som ligger närmare Dummetts användning av indefinit expanderbarhet. Slutligen visar jag att Williamsons argument bäst förstås som ett reductio av legitimiteten av definitionen av det föreslagna predikatet i metaspråket, snarare än antagandet om absolut kvantifikation. Det betyder särskilt att så länge vi inte inför en princip som garanterar definitionens legitimitet så påverkas inte projektet att ge en modellteoretisk semantik för absolut kvantifikation av Williamsons argument. Kapitel 4 vilar på ett gemensamt arbete med Christian Bennet.[4]

I kapitel 5 diskuterar jag tre teorier som på olika sätt tar sig an Williamsons utmaning om att konstruera en strikt adekvat semantik för absolut kvantifikation. Två av teorierna är typteoretiska: både Williamson (2003) och Rayo (2006) använder högre ordningens språk i konstruktionen av semantiken, men de skiljer sig åt i sina analyser av de högre ordningarnas kvantifikatorer. Williamson föreslår att högre ordningens kvantifikatorer ska tolkas som att de kvantifierar över begrepp och Rayo föreslår att de kan tolkas i högre ordningens plurallogik. Både Williamson och Rayo argumenterar för att högre ordningens kvantifikatorer inte nödvändigtvis medför att vi förskriver oss till entiteter utöver de objekt som faller inom räckvidden för första ordningens kvantifikation.

Både Williamsons och Rayos teorier innehåller en oändlig hierarki av högre ordningens språk. Det är möjligt att ge en strikt adekvat semantik för varje nivå i hierarkin, men det finns ingen nivå som tillåter en strikt adekvat semantik för ett språk som innehåller alla nivåer i hierarkin. Det visar sig också att begreppet *strikt adekvans* inte är definierbart från någon nivå i hierarkin, vilket gör det tveksamt om det kan användas för att motivera ett typteoretiskt angreppssätt. Det faktum att vissa tillsynes oproblematiska begrepp inte kan definieras inifrån en typhierarki tillåter konstruktionen av ett dilemma för en typteoretiker som motiverar det typteoretiska angreppssättet med hjälp av Williamsons argument. Dilemmat utgår ifrån en motsägelse som visar att antingen är det typteoretiska angreppssättet i sig motsägelsefullt, eller så är ett tillsynes väldefinierat predikat illegitimt. Men om ett predikat kan sägas vara illegitimt på sådana grunder möjliggörs ett liknande svar på Williamsons argument. Därmed neutraliseras Williamsons argument som skäl att anamma det typteoretiska angreppssättet.

Linnebo (2006) föreslår ett intressant alternativ till det typteoretiska angreppssättet. I avsaknad av en universell mängd, och på grund av de mängd-

[4] Se Bennet och Karlsson (2008).

teoretiska paradoxerna, utgör ZFC ett otillräckligt ramverk för en adekvat semantik för absolut kvantifikation. En naturlig reaktion är därför att ge upp idén om ZFC som metateori. Linnebo argumenterar dock för att en sådan reaktion är alltför drastisk och föreslår att ZFC utökas med en teori om egenskaper. Den resulterande teorin behöver vara tillräckligt stark för att tillåta konstruktionen av en adekvat semantik och samtidigt undvika Williamsons argument på ett sätt som visar att teorin för egenskaper inte är en ad hoc-lösning. Kapitel 5 avslutas med en diskussion av Linnebos idéer.

Efter en kort diskussion om de vanligaste invändningarna utvecklar jag i kapitel 6, delvis inspirerad av Linnebos angreppssätt, den modellteoretiska semantiken för absolut kvantifikation som beskrivits tidigare i denna sammanfattning.