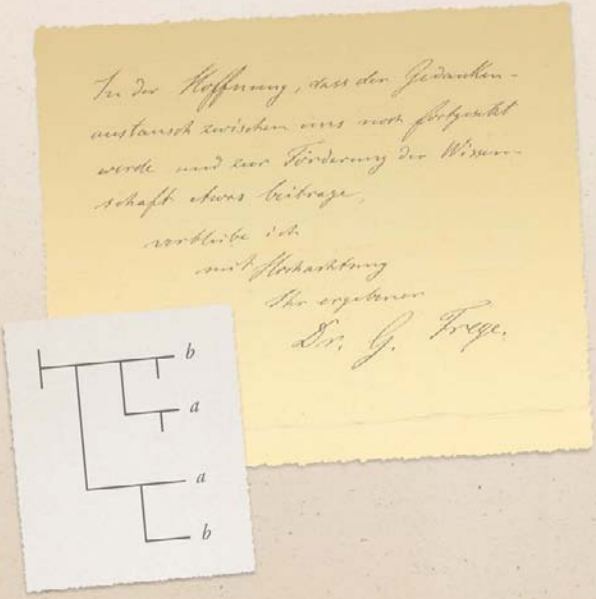
FREGE'S CONCEPTION OF LOGIC



PATRICIA A. BLANCHETTE

Frege's Conception of Logic

This page intentionally left blank

Frege's Conception of Logic

PATRICIA A. BLANCHETTE

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2012 Oxford University Press

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Blanchette, Patricia.
Frege's conception of logic / Patricia A. Blanchette.
p. cm.
Includes bibliographical references.
ISBN 978-0-19-989161-0 (alk. paper)
1. Frege, Gottlob, 1848-1925. 2. Logic. I. Title.
B3245.F24B53 2012
160—dc23 2011029077

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

To Don and Sarah

This page intentionally left blank

TABLE OF CONTENTS

Acknowledgments xi
Note on the Cover xiii
A Note on Citations xv

Introduction 3

1. Logicism and Conceptual Analysis 7
    1.1. INTRODUCTION TO ANALYSIS AND PROOF 7
    1.2. ANALYSIS AND PROOF IN 1879 12
        1.2.1. Derivation, Proof, and Definition 12
        1.2.2. Conceptual Analysis 13
    1.3. ANALYSIS IN 1881 18
    1.4. ANALYSIS IN 1884 19
    1.5. GRUNDGESETZE 22
    1.6. THE GENERAL PICTURE 23

2. Thoughts 28
    2.1. THOUGHTS AND LANGUAGE 28
    2.2. SENSE AND REFERENCE 37
    2.3. THE STRUCTURE OF SENSE 39
    2.4. THOUGHTS AND LANGUAGE AGAIN 45
    2.5. WHERE WE ARE 50

3. Thoughts and Sharp Boundaries 55
    3.1. THE ISSUE 55
    3.2. THE TEXTS 56
    3.3. PIECEMEAL DEFINITION AND NEW OBJECTS 62

3.4. ORDINARY DISCOURSE 68
3.5. CAESAR 70
3.6. QUANTIFICATION 72
3.7. CONCLUSION 75

4. The Analysis of Arithmetic 77
4.1. THE ISSUE 77
4.2. ANALYSIS AS THOUGHT-PRESERVING? 79
4.3. REFERENCE-PRESERVATION AND ANALYSIS 82
4.4. DUMMETT ON WHAT'S PRESERVED 87
4.5. WHAT'S PRESERVED 89
4.5.1. The Case of Directions 89
4.5.2. Numbers 92
4.5.3. Arithmetic 95
4.5.4. Alternative Reductions 101
4.6. CONCLUSIONS 102

5. Analysis and Consistency: The Case of Geometry 108
5.1. INTRODUCTION 108
5.2. FREGE-HILBERT 109
5.3. HILBERT'S METHOD 110
5.4. FREGE'S OBJECTIONS 112
5.5. CONSISTENCY AND CONCEPTS 121
5.6. ANALYSIS AND CONSISTENCY 123
5.7. THE 1906 PASSAGE 125
5.8. ULTIMATE ANALYSES? 128
5.9. CONCLUDING REMARKS 130

6. Frege and Models 135
6.1. MODELS AND CONSISTENCY 135
6.2. MODELS AND ENTAILMENT 137
6.3. IMPLICATIONS 144
6.4. SUMMING UP 146

7. Metatheory 149
7.1. FREGE'S METATHEORY 149
7.2. UNIVERSALISM AND METATHEORY 156
7.2.1. The Issue 156
7.2.2. Internal Tensions 162

7.3. SOUNDNESS, COMPLETENESS, AND CONSISTENCY 165

7.4. CATEGORICITY 167

7.5. CONCLUSION 170

8. Conclusion 175

Bibliography 183

Index 189

This page intentionally left blank

ACKNOWLEDGMENTS

This book has been a long time in the writing, with the result that I am indebted to many, many people and institutions who have supported its writing in a number of different ways. To begin with, many friends and colleagues have taught me a lot about Frege, both in person and through their written work. The influence of that written work will, I hope, be apparent throughout the book. But in case it isn't, I should mention especially the work of Michael Dummett, George Boolos, and Richard Heck as having been especially significant to the development of my thoughts about Frege. Those Frege scholars from whom I've profited additionally via conversation and correspondence include Jamie Tappenden, Michael Kremer, Matthias Schirn, Wilfrid Hodges, Sebastien Gandon, and Joan Weiner. Many students and colleagues at Notre Dame have been helpful in providing feedback, as have audiences to whom some of this material has been presented, especially at Notre Dame, Chicago, Madison, the École Normale Superieur in Paris, and the Ludwigs-Maximilians-Universität in Munich. Special thanks are due to Sebastien Gandon, Marian David, Sean Walsh, and a very helpful anonymous referee for Oxford University Press, all of whom offered comments on draft versions of the manuscript.

Thanks to the University of Notre Dame and to the American Philosophical Society for providing sabbatical support. And thanks to my colleagues in the Philosophy Department at Notre Dame for providing an excellent and congenial environment in which to work.

Finally, for the intangible kinds of support that make everything possible and make it all worthwhile, thanks to my family: my parents, my brothers, the extended tribe, and Ani and Katy. This book is dedicated to Don and Sarah, the center of my universe.

This page intentionally left blank

NOTE ON THE COVER

The cover art is taken from Frege's June 24, 1891 letter to Edmund Husserl. Many thanks to Professors Gottfried Gabriel, Christian Thiel, and Friedrich Kambartel, editors of Frege's *Briefwechsel*, and to Professor Ullrich Melle, director of the Husserl Archives in Leuven, for permission to use the letter.

This page intentionally left blank

A NOTE ON CITATIONS

Frege's main works, *Begriffsschrift, Grundlagen*, and *Grundgesetze I* and *II*, are cited by section number. English quotations are, unless otherwise noted, taken from the standard translations, noted in the Bibliography. Frege's other works are cited by page number in the form "n/m," the first numeral referring to the standard or original German pagination, the second to the standard English translation, both noted in the Bibliography.

This page intentionally left blank

Frege's Conception of Logic

This page intentionally left blank

Introduction

The purpose of this book is to provide a clear account of Frege's conception of logic, with a particular emphasis on those aspects of his conception that diverge from mainstream views of logic today. The reason to pursue the differences that separate modern views from Frege's is in part simply scholarly: we can't get Frege right if we take into account only those of his views that have survived intact. But more importantly, the goal is to take Frege's views seriously on their own terms in an effort to determine to what extent he was right.

Frege's view of logic is closely connected with his conception of, and his commitment to, a certain sort of theoretical reduction. His guiding intellectual project was the attempt to establish his logicist thesis, the thesis that arithmetic is *reducible* to logic. The reduction Frege pursued was of a simple kind, at least in outline: the thesis is that each arithmetical truth is provable from purely logical premises.[1] This immediately puts logic at center stage in two ways: one needs an identifiable collection of logical truths to stand as premises, and a reliable means of employing logic in the course of proof.

There is little that is controversial about the general idea that to *prove* the claims of one theory using just the resources of another is to effect an important reduction, one that shows that the reduced science is grounded, in an explanatorily useful way, in the reducing one. As Frege puts it, the effect of such a reduction of arithmetic to logic would be to show that arithmetic is "simply a development of logic, and every proposition of arithmetic a law of logic, albeit a derivative one."[2] But considerably less straightforward is Frege's idea that his way of proceeding would in fact have constituted a demonstration that arithmetical truths are provable from logical ones. Leaving aside the infamous difficulty about his purportedly logical truths, i.e. that some of them are now known not to be logical and arguably not even to be truths, a further difficulty surrounds the purportedly arithmetical ones. The truths Frege actually proves are not easily-recognizable claims about e.g. the commutativity of addition or about the number 7. They are instead truths about complex and apparently-unfamiliar objects, functions, and relations. The immediate question to press on Frege is that of how the provability

from logical premises of these newly-displayed and unfamiliar claims is supposed to demonstrate the provability from logical premises of the truths of *arithmetic*.

Frege's answer to this question is one that we need to patch together from his practice and from a number of his remarks in the vicinity of the topic. The general idea, to be pursued at length in what follows, is that the truths Frege proves are related to the truths of arithmetic via conceptual analysis. And this is where Frege's conception of logic and his conception of theoretical reduction have their most interesting convergence. The relation of logical entailment, Frege seems to think, survives conceptual analysis: if a collection of premises logically entails a conclusion *c*, then it logically entails the pre-analytic version of that conclusion, the claim that, when subjected to conceptual analysis, yields *c*.[3] This general idea is familiar and uncontroversial in easy cases: if, for example, knowledge were in fact appropriately analyzed as justified true belief, then we would all presumably agree that any premises entailing that Sarah has a justified true belief that *p* would suffice to entail that Sarah knows that *p*. But as we'll see, the thesis itself is considerably more interesting than its easy instances reveal.

Frege's view of logical entailment as surviving conceptual analysis is, as just noted, essential to his reductionist project. It is also central to his conception of logic in general, and especially to some of the most interesting ways in which that conception diverges from its currently-mainstream descendants. The utility of conceptual analysis as a way of revealing logical connections underlies a deep difference between Fregean and more-modern ways of understanding the connection between languages and the kinds of things expressed by the sentences or formulas of those languages. As a consequence, it leads to important differences between Frege and his successors over the role and importance of formal derivations, particularly with respect to the connection between a derivation of a sentence or formula and the proof of the claim expressed by that sentence or formula. This difference gives rise to a difference over the implications for a formal theory of the existence of a model of that theory. In particular, the use of models to demonstrate consistency results, as is now the universal practice, is seriously problematic from Frege's point of view. The Fregean conception of logic also gives rise, for similar reasons, to what are by today's lights some nonstandard views regarding ordinary results in metatheory. A large part of the point of this book is to clarify and make sense of the standpoint from which these unusual views seem natural.

The plan of the book is as follows.

We begin by taking a careful look at Frege's logicist reduction, particularly at the role of conceptual analysis and its connection with logic in that project. Chapters 1–3 explore the connection between logic and conceptual analysis, and investigate the role played by Fregean thoughts in this connection. Thoughts, i.e. those nonlinguistic propositions expressed by declarative sentences, are for Frege the substance of mathematical and scientific theories. They are also the

bearers of such logical relations as entailment and provability, and the relata of conceptual analysis. Hence it's important to get clear about their role in Frege's reductionist project if we are to understand his general conception of logic. It will be argued here that attention to the role of thoughts in Frege's analytic project will make it clear that his overall reductionist program is considerably more closely tied to providing an account of the content of ordinary arithmetical discourse than is often supposed. Chapter 4 substantiates this reading of Frege with a careful look at the analytic work of *Grundlagen* and *Grundgesetze*.

Frege holds that the relation of logical entailment between thoughts can often be discovered only after a good deal of nontrivial conceptual analysis as applied to those thoughts. As just noted, this view has implications for the role of formal derivations in establishing logical entailment, and for the role of models in establishing consistency and independence. Chapters 5–7 explore some of the central such implications. It's argued in Chapter 5 that Frege's rejection of Hilbert's 1900 independence and consistency proofs in the latter's *Foundations of Geometry* is well motivated, given Frege's understanding of independence and consistency. This, as explored in Chapter 6, means that both the modern technique of using models to establish consistency, and the use of a Tarskian account of model-theoretic entailment as an indication of logical entailment, are importantly flawed from Frege's point of view. The upshot of this discussion is not that these standard modern techniques are problematic but rather that there are two quite different ways of understanding the relation of logical entailment and its expression in formal systems—the Fregean and the post-Fregean—and that the modern techniques in question are successful with respect only to the latter such understanding. The reasonableness of Frege's own understanding of logical entailment, as is argued in Chapters 3 and 4, provides reason for care when taking modern results to have implications relevant to pretheoretic conceptions of e.g. entailment and consistency. The difference between Fregean and subsequent accounts of the logical relations, in any case, serves as a reminder against reading Frege too easily through the lens of later work.

But there are also, I'll argue, ways in which Frege's distance from his successors has been overstated in recent scholarship. Chapter 7 takes up the question of whether Frege did, and of whether a Fregean can reasonably, engage in significant metatheoretical work. The answer here is "yes." Metatheory looks in several interesting ways quite different from Frege's point of view than it does from, say, Tarski's; but Frege's work is decisively metatheoretic in straightforward ways, and much of modern metatheory can unproblematically be made sense of from a purely Fregean perspective. The philosophical implications of specific metatheoretic results, on the other hand, often appear quite different from a Fregean than from a later point of view.

The picture to be presented here is one according to which much of what now seems odd or mistaken about Frege's conception of logic is both natural and plausible when understood against the background of his own attractive picture

of the connection between conceptual analysis and logic. My hope is that an appreciation of this point can lead not just to a better understanding of Frege but also to a clearer appreciation of the underpinnings of the traditions that succeed and build upon his work.

Notes

1. Frege often puts the point in terms of the provability of arithmetical truths from laws of logic *and definitions*. As we'll see in Chapter 1, the eliminability of the definitions entails that the proofs make essential appeal only to the principles of logic.
2. *Grundlagen* (hereafter *GI*) §87.
3. Strictly speaking, then, the idea here is that logical entailment survives the converse of conceptual analysis. As we'll see, it goes both ways: if *c* is the result of a successful conceptual analysis of *c′*, then c and *c′* are logically entailed by the same premises.

1

Logicism and Conceptual Analysis

1.1 Introduction to Analysis and Proof

The central intellectual project of Frege's life was the development and defense of his own particular brand of logicism, that is, of the thesis that arithmetic is grounded in logic. The project occupies him from 1879, when he introduces the topic in the *Begriffsschrift* monograph, until 1902, when he learns from Russell that his approach is fundamentally flawed. In pursuit of the project, Frege develops a lovely formal system for quantified logic; a detailed analysis of fundamental arithmetical concepts, objects, and truths; and a series of painstakingly rigorous proofs of the truths so analyzed from what he took to be basic truths of logic. In this chapter, we'll sketch the outlines of the logicist project as it develops from 1879 to 1903, with an emphasis on coming to understand the role of conceptual analysis in that project.

The logicist question is first raised in *Begriffsschrift* as follows:

> The most reliable way of carrying out a proof, obviously, is to follow pure logic, a way that, disregarding the particular characteristics of objects, depends solely on those laws upon which all knowledge rests. Accordingly, we divide all truths that require justification into two kinds, those for which the proof can be carried out purely by means of logic and those for which it must be supported by facts of experience. . . . Now, when I came to consider the question to which of these two kinds the judgments of arithmetic belong, I first had to ascertain how far one could proceed in arithmetic by means of inferences alone, with the sole support of those laws of thought that transcend all particulars. My initial step was to attempt to reduce the concept of ordering in a sequence to that of *logical* ordering, so as to proceed from there to the concept of number. To prevent anything intuitive from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free of gaps. In attempting to comply with this requirement in the strictest possible way I found the inadequacy of language to be an obstacle; no

> matter how unwieldy the expressions I was ready to accept, I was less and less able, as the relations became more and more complex, to attain the precision that my purpose required. This deficiency led me to the idea of the present concept-script. Its purpose, therefore, is to provide us with the most reliable test of the validity of a chain of inferences and to point out every presupposition that tries to sneak in unnoticed, so that its origin can be investigated.[1]

Three points worth remarking about this passage, for our purposes, are as follows. First, the logicist question as it's characterized here is immediately a proof-theoretic one: the question is that of whether the truths of arithmetic are *provable* "purely by means of logic." This way of understanding the issue remains constant throughout Frege's logicist period. Second, Frege notes here that the "initial step" in the proof-theoretic project is one of conceptual analysis: he needs to "reduce the concept of ordering in a sequence to that of logical ordering" before giving the important proofs. We'll look in some detail in this chapter (§1.2) at this reduction and its connection with proof. For now, we'll simply note that this general idea—that a certain kind of conceptual reduction or analysis is sometimes required as a preliminary to proof—is a pervasive theme throughout all of Frege's logicist works. Finally, we have in this passage a clear statement of the purpose of Frege's concept-script, i.e. of his formal system. The rationale for its introduction is simply that of rigor: it is designed for the presentation of perspicuous and gap-free proofs.

Much of the *Begriffsschrift* monograph is taken up with the work of introducing the formal system. In the third part, Frege offers the promised reduction of what he calls "ordering in a sequence," the central piece of which is the account of the ancestral of an arbitrary binary relation. We'll take a close look at this reduction in §1.22. Having given the reduction, Frege next provides rigorous formal proofs for a handful of important claims about sequential ordering, appealing in each case only to principles of pure logic.[2]

Despite these successful initial steps of analysis and proof, Frege doesn't take a stand in *Begriffsschrift* on the general logicist question. He sums up his accomplishments and future plans at the end of *Begriffsschrift*'s preface as follows:

> As I remarked at the beginning, arithmetic was the point of departure for the train of thought that led me to my concept-script. And that is why I intend to apply it first of all to that science, attempting to provide a more detailed analysis [or decomposition: *zergliedern*] of the concepts of arithmetic and a deeper foundation for its theorems. For the present I have reported in the third Part some of the developments in this direction. To proceed further along the path indicated, to elucidate the concepts of number, magnitude, and so forth—all this will be the object of further investigations, which I shall publish immediately after this booklet.[3]

The next stage in the "analysis" or "elucidation" of arithmetical concepts appears in Frege's 1884 *Grundlagen*. Here again, Frege is primarily concerned with the question of how, and on what basis, one might prove the truths of arithmetic. And, as with the proof-theoretic project of *Begriffsschrift*, the strategy involves some crucial preliminary conceptual analysis. As he puts it in *Grundlagen*'s early pages:

> [T]he fundamental propositions of arithmetic should be proved, if in any way possible, with the utmost rigor; for only if every gap in the chain of deductions is eliminated with the greatest care can we say with certainty upon what primitive truths the proof depends. . . .
>
> If we now try to meet this demand, we very soon come to propositions which cannot be proved so long as we do not succeed in analyzing [or "resolving": *aufzulösen*] concepts which occur in them into simpler concepts or in reducing them to something of greater generality. Now here it is above all cardinal number which has to be either defined or recognized as indefinable. This is the point which the present work is meant to settle. On the outcome of this task will depend the decision as to the nature of the laws of arithmetic.[4]

As in *Begriffsschrift*, the idea here is that the proof of arithmetical truths somehow waits on our success in "analyzing concepts which occur in them into simpler concepts or in reducing them to something of greater generality." If we are to take Frege at his word, the possibility of providing such an analysis or reduction of the concept of cardinal number is "*the point* which [*Grundlagen*] is meant to settle."[5]

Much of *Grundlagen* is taken up with arguments concerning the nature of arithmetical truth, a good deal of which takes the form of criticism of competing views. Arithmetical truths cannot, Frege argues, be understood as empirical generalizations or as psychological reports. Nor can they be understood as truths about instants of time or portions of space. All such suggestions, Frege argues, badly misconstrue the content of ordinary arithmetical discourse. He argues that numbers (*Anzahlen*) are fundamentally measures of cardinality, in the sense that the "basic use" of number words is in assignments of cardinal size, i.e. in statements of the form "there are *n* F's."[6] This argument concludes, via steps we'll examine in Chapter 4, with Frege's infamous account of the numbers as the extensions of a particular series of concepts.

Having given the analyses, Frege provides in *Grundlagen* proof-sketches for some fundamental claims about cardinal number, showing in outline how those propositions, understood in terms of the just-delivered conceptual analyses, might be demonstrated from purely logical premises. As Frege points out here, the case for logicism will not be sealed until the proof-sketches are replaced by gap-free and rigorous proofs:

> [I]t can still always be doubted whether [arithmetical propositions] are deducible solely from purely logical laws, or whether some other type of premise is not involved at some point in their proof without our noticing it. This misgiving will not be completely allayed even by the indications I have given of the proof of some of the propositions; it can only be removed by producing a chain of deductions with no link missing, such that no step in it is taken which does not conform to some one of a small number of principles of inference recognized as purely logical.[7]

The provision of just such rigorous proofs is the purpose of Frege's [1893/1903] *Grundgesetze*.

The ratio of conceptual analysis and defense thereof to proof (or proof-sketch) is very high in *Grundlagen*. That early work is dedicated as Frege says to a discussion of the question of how best to analyze number-theoretic discourse, and contains only a very brief introduction to the proof-theoretic project proper. The balance is reversed in *Grundgesetze*: here the goal is to give the all-important proofs, relying on the earlier work for much of the analytic part. Frege claims, though we will need to investigate to what extent he should be taken seriously here, that the work of *Grundgesetze* "rests on" the analysis provided in *Grundlagen*.[8] *Grundgesetze* does offer some of its own analytic work, including some interesting revamping, discussed below, of some of the territory already covered in *Grundlagen*, and a significant expansion of the arithmetical domain covered. But, in brief, the work of *Grundgesetze* is primarily that of proof, while *Grundlagen*'s is primarily that of spadework prior to proof.

Though his attention is focused more on proof and less on conceptual analysis after 1884, there is a good deal of evidence that Frege maintains throughout his lifetime the general picture sketched in *Grundlagen*, namely, that the careful analysis or elucidation of arithmetical truths and their components can play a crucial preparatory role with respect to proof. His discussion of the nature of analysis changes over the years, with the largest change owing to the maturing of his semantic theory in the early 1890s. Prior to 1891, Frege characterizes arithmetical sentences and their constituent words as expressing *contents*, some of which he rather vaguely refers to as "concepts." As we have seen, his discussions of analysis and elucidation in this period typically take those concepts to be the target of conceptual analysis. After 1891 and the advent of Frege's famous two-tiered semantic theory, arithmetical sentences and their constituent words each have both a sense and a reference; discussions of conceptual analysis and elucidation after this point typically have it that the targets of conceptual analysis are *senses*. The term "concept" ("*Begriff*") itself takes on a narrow technical meaning after 1890, and Frege avoids its use in characterizing the targets of elucidation. We examine these and related changes in more detail in Chapter 2; the central point for present purposes is that

throughout Frege's work we find the same core idea: that it's possible to demonstrate the ultimate grounds of a body of truths via a two-step process involving first a careful elucidation or analysis and second the rigorous proof of the thus-analyzed truths from the simplest possible premises. As Frege characterizes the role of analysis in proof as late as 1914, echoing the sentiment of 1884:

> In the development of a science it can indeed happen that one has used a word, a sign, an expression, over a long period under the impression that its sense is simple until one succeeds in analyzing it into simpler logical constituents. By means of such an analysis, we may hope to reduce the number of axioms; for it may not be possible to prove a truth containing a complex constituent so long as that constituent remains unanalyzed; but it may be possible, given an analysis, to prove it from truths in which the elements of the analysis occur.[9]

This still-vague account gives rise to a number of questions, particularly when we look at the details of its implementation in Frege's work. The first question is just that of what Frege means by "analysis" or "elucidation." The second concerns the plausibility of his appeal to analysis in the overall project: one wants to know whether it is in fact reasonable to take the provability of an analysans proposition from some premises P as a guarantee that the original analysandum proposition itself is provable from P. Does Frege's conception of analysis underwrite his crucial inference from the provability (from logical premises) of his own highly-analyzed versions of arithmetical truths to the conclusion that the arithmetical truths themselves are so provable? As we'll see, the distance between Frege's analysantia and their original analysanda is sufficiently great as to make it far from obvious that this inference is warranted. And finally, this distance itself and the consequent difficulty for the central Fregean inference provide reason to doubt the account of Frege's work just presented: it has been argued that Frege's proofs cannot be understood to be proofs of analyzed versions of arithmetical truths, but must instead be read as proofs of the truths of a different science altogether.[10]

The line of interpretation suggested in the last sentence will be resisted in what follows: I take it that Frege was in fact interested in discovering the grounds of arithmetic and not of some newly created science. But the pressure to turn to such a reading is real, owing to the highly nontrivial nature of Frege's analyses. This makes the first two questions noted above particularly pressing: what does Frege take an analysis or elucidation to be, and is his employment of these things in his proof-theoretic project justified? The first stage of answering these questions is to examine in some detail a sample of Frege's important analyses. We turn to these, in chronological order, next.

1.2 Analysis and Proof in 1879

1.2.1 Derivation, Proof, and Definition

The central work of Frege's *Begriffsschrift* is the introduction of his formal system and the application of that system to the proof of some fundamental claims about, as he puts it, "ordering in a sequence." The system itself, which we'll call "*B*," consists of a formal language, axioms for higher-order quantified logic, and the rule of inference *modus ponens*.[11] *B* is "formal" in the sense that its well-formed formulas, its axioms, and its inference-rule are all defined syntactically: the question of whether a given series of marks on paper counts as a legitimate derivation within the system is a matter to be determined entirely in terms of the shapes of those marks. But Frege's language is not "formal" in the sense of "uninterpreted." His intention is that each formula in a derivation will express a determinate proposition, referred to by Frege variously as a "content of possible judgment," a "conceptual content," or simply a "content."[12] The central idea is that a series of formulas sanctioned by *B*'s rules will express a series of contents, the last of which is thereby *proven* from the contents expressed by premise-formulas, if any.

It will be important in what follows to distinguish clearly a series of formulas sanctioned by a formal system from the series of contents expressed by those formulas. We'll refer to the former as a *derivation*, and to the latter as a *proof*. The last line of a derivation expresses the judgment-content proven, and the proof expressed by that derivation establishes that its final judgment-content follows logically from the premises, if any, assumed in the proof. A derivation that contains no premise-formulas establishes that the judgment-content expressed by its last line is a truth of logic. As we've seen in the passages quoted above, Frege's purpose in using the concept-script, rather than German, is that of maximizing precision, clarity, and rigor. A central demand to be met by the formal system is that no presupposition can "sneak in unnoticed" in the course of a derivation.[13] And the guarantee of this is that each step of a derivation must meet the strict rules laid down in advance, the clarity and rigor of which are ensured by their syntactic specification.

Before turning to the details of his analyses, we need to briefly lay out Frege's straightforward understanding of the role of definitions within derivations. A definition that appears within a Fregean derivation is always a stipulation of notational convention. It is an announcement that a new piece of notation is to be used as shorthand for an already-understood complex of terms. As Frege puts it immediately after the first occurrence of the *Begriffsschrift*'s first definition, formula (69):

> This sentence differs from the judgments considered up to now in that it contains signs that have not been defined before; it itself gives the definition. It does not say "The right side of the equation has the same content as the left," but "It is to have the same content."

> . . . [W]e can do without the notation introduced by this sentence [(69)] and hence without the sentence itself as its definition; nothing follows from the sentence that could not also be inferred without it. Our sole purpose in introducing such definitions is to bring about an extrinsic simplification by stipulating an abbreviation.[14]

If not for the fact that extremely long formulas are hard to work with, there would be no role for definitions in Frege's derivations. The new terms they introduce express exactly what's expressed by the complexes of terms they abbreviate. Hence it's important to keep in mind that nothing like a "conceptual analysis" is involved in Frege's formal definitions, i.e. in those definitions that appear in the course of formal derivations. There is no sense in which the definiendum or what it expresses is being reduced, or analyzed, in terms of the definiens or what it expresses; all we have in such a definition is notational simplification.

It is crucial, given the purpose of Frege's derivations, that definitions appearing in them introduce no hidden assumptions or other nonlogical content. Frege's guarantee of this is the conservativeness point noted above, that "nothing follows from [a definition] that could not also be inferred without it." Here the point is not of course that no new *formulas* are derivable in *B* when a definition is added; derivable formulas involving the newly introduced term are newly derivable in this sense. The point is rather that no new judgment-contents are provable via the use of definitions unless provable without.[15] For every formula containing a defined term expresses precisely the judgment-content expressed by its long-winded counterpart, i.e. by the formula obtained from it when we replace all defined terms by their definientia, repeatedly, until arriving at a sentence in primitive terminology. Every axiom containing a defined term remains an axiom if that term is replaced by its definiens, and every instance of a rule of inference remains such an instance if defined terms are replaced by definientia. Hence any derivation involving a definition can be straightforwardly transformed into a derivation that includes no definitions, and whose premise-formulas and concluding formula are simply disabbreviated versions of the original premise-formula and concluding-formula, respectively. The conservative role of definitions justifies Frege's later claim that if his conclusions are provable using just laws of logic and definitions, they will have been shown to be provable from, and hence grounded in, pure logic.

1.2.2 Conceptual Analysis

Conceptual analysis first appears in Part III of *Begriffsschrift*, in which Frege provides the promised "reduction" of the relation of *ordering in a sequence* and of subsidiary notions. Part III opens as follows:

> The derivations that follow are intended to give a general idea of the way in which our concept-script is handled. . . . Through the present example, moreover, we see how pure thought, irrespective of any content given by the senses or even by an intuition a priori, can, solely from the content that results from its own constitution, bring forth judgments that at first sight appear to be possible only on the basis of some intuition.[16]

That is to say, a handful of judgment-contents that look "at first sight" as if they would be knowable only via sensation or a priori intuition will be shown here to be in principle knowable without appeal to intuition. The demonstration that they are in principle so knowable—i.e., that they are grounded in pure logic—consists of a proof from principles of logic and definitions of analyzed versions of the judgment-contents in question.

The central notion involved in the Part III analyses is that of the strong ancestral of an arbitrary binary relation *f*, or, as Frege calls it, that of "following in the *f*-sequence." An object *b* bears this relation to object *a* iff *b* can be obtained from *a* by a finite number of iterations of *f*. That is, *b* follows *a* in the *f*-sequence iff either $f(a, b)$, or for some intervening *x*: $f(a, x)$ and $f(x, b)$, or finally, for some finite intervening collection $x_1 \ldots x_n$: $f(a, x_1)$ and . . . and $f(x_{n-1}, x_n)$ and $f(x_n, b)$. When $f(x, y)$ is the relation $(y = x + 1)$, the relation of *following in the f-sequence* as applied to integers is just the greater-than relation; the numbers that follow 0 in this relation are the positive integers greater than zero. A good account of the relation of following in the *f*-sequence will eliminate the appeal to finitude in the gloss just given.

Frege's strategy involves, first, a few definitions, the two most important of which are (69) and (76). Definition (69) introduces terminology—which we have abbreviated below as "Her(*F*, *f*)"—for what Frege calls "the circumstance that property *F* is hereditary in the *f*-sequence," for *F* any unary and *f* any binary relation. Frege's definition can be put into modern terminology as follows, with "*F*" and "*f*" schematic:

$$(69) \quad \vdash\!\!\!\vdash \quad (x)(Fx \rightarrow (y)(fxy \rightarrow Fy)) \equiv \mathrm{Her}(F,f)^{17}$$

Henceforth, substituting for "*F*" and "*f*" any one-place first-level predicate and two-place first-level predicate respectively, the formula on the right-hand side of (69), so substituted, is to be treated as an abbreviation of its counterpart on the left.[18]

Definition (76) employs the symbols defined in (69) and introduces terminology that we have abbreviated as "$Fol_f(y, x)$". The definition is, with "*f*" schematic:

$$(76) \quad \vdash\!\!\!\vdash \quad (\mathrm{F})(\mathrm{Her}(F,f) \rightarrow ((z)(fxz \rightarrow Fz) \rightarrow Fy)) \equiv Fol_f(y,x)^{19}$$

In short, "$Fol_f(y,x)$" says that y has all those properties that are both hereditary in the *f*-sequence and had by every *f*-successor of *x*.[20]

Among the formulas subsequently derived using just definitions and principles of logic is the especially interesting.

$$(98)\ \vdash\quad Fol_f(y,x) \rightarrow (Fol_f(z,y) \rightarrow Fol_f(z,x))$$

which says that, for arbitrary *f*, the relation $Fol_f(y,x)$ is transitive. And here, of course, since "$Fol_f(y,x)$" is an abbreviation, (98) is shorthand for the disabbreviated

$$\begin{aligned}(98^*)\ \{&(F)((x)(Fx \rightarrow (y)(fxy \rightarrow Fy)) \rightarrow ((w)(fxw \rightarrow Fw) \rightarrow Fy)) \rightarrow \\ &[(F)((x)(Fx \rightarrow (y)(fxy \rightarrow Fy)) \rightarrow ((w)(fyw \rightarrow Fw) \rightarrow Fz)) \rightarrow \\ &(F)((x)(Fx \rightarrow (y)(fxy \rightarrow Fy)) \rightarrow ((w)(fxw \rightarrow Fw) \rightarrow Fz))]\}.\end{aligned}$$

(98*) expresses the transitivity of the relation $Fol_f(y,x)$, i.e. the relation borne by y to *x* iff *y* has all of those properties that are hereditary in the *f*-sequence and had by all of *x*'s *f*-successors. The derivation of (98) demonstrates that the transitivity of this relation is grounded in pure logic.

Frege appends to (98) a translation into ordinary language as follows:

> (TRANS) If *y* follows *x* in the *f*-sequence and if *z* follows *y* in the *f*-sequence, then *z* follows *x* in the *f*-sequence.[21]

The relation referred to in (TRANS) as "following in the *f*-sequence" is presumably the relation with which Frege's readers can be assumed to be familiar, the relation that is straightforwardly cashed out in terms of *y*'s being obtainable via a finite number of applications of *f* to *x*. As Frege helpfully notes on introducing the term:

> To make clearer the generality of the concept, given hereby, of succession in a sequence, I remind the reader of a number of possibilities. Not only juxtaposition, such as pearls on a string exhibit, is subsumed here, but also branching like that of a family tree, merging of several branches, and ringlike self-linking.[22]

The transitivity of the familiar relation is one of those that, as Frege suggests, might very naturally have struck the uninitiated as dependent on e.g. the intuitively given ordering of space or of time. Thinking of pearls on a string, the fact that: if *y* can be obtained from *x* by a finite number of applications of the function *one pearl to the right*, and *z* can be so obtained from *y*, then *z* can be so obtained from *x*, may indeed seem to be dependent on nontrivial facts about the ordering of points in space.

Frege's derivation of (98) is supposed to convince his audience that this is not the case, i.e. that the transitivity of this kind of ordering is a matter of pure logic. But how, exactly, is it supposed to do this? The derivation of (98), recall, immediately demonstrates that the transitivity of the relation $Fol_f(y,x)$, i.e. of the relation that holds when y has all of the *f*-hereditary properties of *x*'s *f*-successors, is grounded in pure logic. What we want to know is whether the transitivity of the familiar relation of *following in the f-sequence*—that relation that's explicable in terms of the finite iteration of *f*—is so grounded.

The answer that we can give by simply adopting rather than explaining Frege's terminology is that the latter relation, the one with which his audience would have been familiar, has been "analyzed" or "reduced" in terms of the former. The fact, or the claim, that *b* can be obtained from *a* by a finite number of applications of *f* is, on this picture, explicated in terms of *b*'s possession of all those *f*-hereditary properties had by *a*'s *f*-successors. As Frege says, he has "reduce[d] the concept of ordering in a sequence to that of *logical* ordering. . . ."[23]

To say this is not to do much more than to label the important connection between the ordinary and the "reduced" claim. What one wants to know is how, exactly, Frege conceives of the relationship between that ordinary claim, the one typically expressed by (TRANS), and its highly-analyzed counterpart, the one expressed by (98). Unfortunately, Frege doesn't devote any attention in 1879 to this further question; his project here involves the exhibition of what he takes to be good analyses, rather than an account of what a good analysis consists in.

In this particular case, it's hard to argue with Frege's presumption that the reduction is a reasonable one. The difference between the content of the ordinary

(α) *y* follows *x* in the *f*-sequence

and Frege's

(α^*) *y* has every property that is both hereditary in the *f*-sequence and is had by *x*'s *f*-successors

is not large. It would be unusual to challenge Frege's claim to have proven (TRANS) on the grounds that his proven proposition is properly about the transitivity of the relation expressed by (α^*) rather than of that expressed by (α).[24] Whatever, exactly, a conceptual analysis is intended to provide, both intuition and subsequent mathematical practice support Frege's idea that it is provided here: the derivation of (98) from purely logical principles shows, as Frege says, that (TRANS) is grounded in pure logic. Hence we have in the derivation of (98) an instance of the anti-Kantian demonstration that Frege promises at the outset of Part III.[25]

But it is important to note that the justification for Frege's claim to have demonstrated the purely-logical grounding of (TRANS) turns on the adequacy of the analysis of (α) in terms of (α*). This is not because the cogency or the rigor of the proof itself rests on a claim of analytic adequacy; the proof stands on its own. Frege's derivation of formula (98) clearly demonstrates the purely-logical grounding of the content expressed by (98), i.e. of the content expressed by (98*). That the analysis of (α) in terms of (α*) is a *good* analysis is what underwrites Frege's claim, on the basis of a derivation of (98), to have established the purely-logical grounds of the familiar claim that *following in the f-sequence* is transitive.

This pattern is repeated as the logicist project develops. Each Fregean derivation, and hence each Fregean proof, stands on its own: the proof demonstrates that its conclusion follows purely logically from its premises, and hence that in the case of premise-free proofs, the conclusion is a truth of logic. This demonstration turns not at all on Frege's conceptual analyses. But that the purely-logical status of the formally proven conclusion establishes what Frege really wants, i.e. the purely-logical status of a closely connected truth about ordering, or about arithmetical objects and functions, requires that Frege's proposed analyses are accurate. It is essential to the work to which Frege puts his proofs in the defense of logicism that his proven analysantia are sufficiently similar to their counterpart analysanda that the purely-logical status of the former suffices for that of the latter. To look ahead for a moment: when Frege proves in *Grundgesetze* his highly-complex claims about extensions of concepts, thereby establishing the purely-logical status of these unfamiliar claims, the further Fregean thesis that these proofs establish the purely-logical grounds of *arithmetic* requires that Frege is right to propose the unfamiliar, proven claims as appropriately faithful analyses of their ordinary counterparts, the truths of arithmetic.

The remainder of *Begriffsschrift* includes a handful of similar proofs, demonstrating that several central claims about series are in fact (assuming the adequacy of Frege's analyses) grounded in pure logic.

Recall that Frege closes the preface to *Begriffsschrift* with the following promise:

> To proceed further along the path indicated, to elucidate the concepts of number, magnitude, and so forth—all this will be the object of further investigations, which I shall publish immediately after this booklet.[26]

The promised "elucidations" of number etc. are carried out in the 1884 *Grundlagen*. Before turning to *Grundlagen*, we pause to look at a manuscript from 1881 that illustrates an interesting stage in the development of Frege's analyses.

1.3 Analysis in 1881

In a manuscript titled "Boole's Logical Calculus and the Concept-Script," submitted by Frege in 1881 (three times, unsuccessfully) for publication, we find Frege employing a clear and deliberative application of his general analytic/proof-theoretic method. In this case, the proof is intended to demonstrate not the logical grounding of an arithmetical truth, but the preliminary result that a given relatively-complex arithmetical truth is grounded in illuminatingly simpler ones. The relevant passage begins (with our labels inserted for later reference):

> I wish to prove the theorem that
>
> (SUM) the sum of two multiples of a number is in its turn a multiple of that number.
>
> . . .
>
> The numbers whose multiples are to be considered are subject to no conditions other than that the following addition theorems:
>
> (P1) $(\forall m)(\forall n)(\forall p)((m + n) + p = m + (n + p))$
>
> and
>
> (P2) $(\forall n)(n = n + 0)$
>
> hold for them.[27]

As Frege notes, the desired result is an interesting one, since the premises (P1) and (P2) are, on the surface at least, surprisingly weak as compared with the conclusion (SUM) to be demonstrated. Having announced the project, Frege gives his reader a careful analysis of the relation "multiple-of," to the effect that a number n is a multiple of a iff n follows 0 in the "$+a$" sequence, with "following in the . . . sequence" further cashed out as in *Begriffsschrift*, and "0" and "+" left unanalyzed. We thereby obtain a highly-analyzed version of what is expressed by (SUM), presented via a complex formula which we'll abbreviate as

(SUM*) $(\forall m)(\forall n)(\forall a)((Fol_{+a}(m,0) \,\&\, Fol_{+a}(n,0)) \rightarrow Fol_{+a}((m+n),0))$[28]

That is, if m follows 0 in the "$+a$"-sequence, and n follows 0 in that sequence, then $(m + n)$ does too. (SUM*) is then the sentence that Frege derives from the premises (P1) and (P2) with his usual painstaking rigor. On the basis of this derivation of (SUM*), Frege takes himself to have established that the theorem expressed by (SUM) follows logically from those premises.

The pattern of analysis and proof exhibited here in 1881 is just as in *Begriffsschrift*, and might be characterized as follows. Beginning with a target judgment-content τ whose grounds are to be investigated, Frege begins with a thorough analysis of τ, typically, as here, via an analysis of its important components, yielding the highly analyzed content τ^*. He then provides a rigorous proof of τ^*,

which proof is intended, if we are to take Frege's commentary at face value, to establish the logical grounds of τ itself. The crucial assumption, again taking Frege at his word, is that τ and τ^* are sufficiently similar that the provability of the latter from particular premises (or from no premises at all) suffices to establish the grounding of the former in those premises (or in pure logic).

This assumption, and the inferences it licenses, are arguably unproblematic in cases like those above from 1879 and 1881, in which the fully analyzed content will strike the ordinary reader as simply a clarification of its original. But things are considerably less simple, and arguably more problematic, when we turn to Frege's more interesting analyses.

1.4 Analysis in 1884

As we've already seen, Frege portrays the work of conceptual elucidation, in *Grundlagen*, as crucial to the proof-theoretic project. One potentially confusing aspect of the *Grundlagen* discussion concerns the use here of the term "definition" ("*Definition*"). To repeat a passage quoted a few pages ago, Frege notes that in the attempt to prove the truths of arithmetic:

> [W]e very soon come to propositions which cannot be proved so long as we do not succeed in analyzing concepts which occur in them into simpler concepts or in reducing them to something of greater generality. Now here it is above all cardinal number which has to be either defined or recognized as indefinable [. . . *definirt oder als undefinirbar anerkannt werden muss*]. This is the point which the present work is meant to settle. On the outcome of this task will depend the decision as to the nature of the laws of arithmetic.[29]

Frege's talk in this passage and in related contexts of "defining" concepts (and objects) needs to be kept clearly distinguished from the stipulative definitions discussed above, those that occur in derivations. Stipulative definitions are, again, simply devices of abbreviation, and do not express the results of conceptual analysis (or anything else, for that matter). They are entirely eliminable. "Definition," in the sense used in this *Grundlagen* passage, on the other hand, refers to the crucial Fregean activity of conceptual analysis. As Frege himself notes, the dual use of the term is unfortunate (though common).[30] The important point here is that the *definition* of the concept of cardinal number that Frege talks about in this *Grundlagen* passage is the analysis of that concept into simpler components.

The central analyses offered in *Grundlagen* are the well-known ones of cardinal number, of zero, of successor, and of finite cardinal. A few of the details are

as follows. Where a first-level concept is a concept under which objects fall, two first-level concepts are on Frege's account *equinumerous* iff there's a 1–1 function mapping the objects falling under one of them onto the objects falling under the other. A second-level concept is one under (or, as he later puts it, "within") which first-level concepts fall. Say that a second-level concept *C* is a *cardinality* concept if for some first-level concept *F*, the concepts falling under *C* are all and only those first-level concepts equinumerous with *F*. Frege takes the cardinal numbers to be the extensions of cardinality concepts. Zero is the extension of that cardinality concept C_0 under which fall all and only those first-level concepts equinumerous with the concept *not self-identical*. (This latter is just the first-level concept under which objects fall if they're not self-identical; hence this concept is guaranteed via pure logic to be empty.) The cardinal number one is the extension of that cardinality-concept C_1 under which fall all and only those first-level concepts equinumerous with the concept *identical with 0*. (Note that this latter concept is guaranteed, via pure logic, to have exactly one object falling under it.) "The extension of the second-level concept *equinumerous with F*," for *F* a first-level concept, is abbreviated as "the number that belongs to the concept *F*." Successor is taken to be that relation that holds between *n* and *m* iff for some concept *F* and some object *x* falling under *F*, *n* is the number that belongs to *F*, and *m* is the number that belongs to the concept *falling under F, but not x*. The finite cardinals are 0 together with those things that follow 0 in the successor-sequence, where "follows . . . in the . . . sequence" is understood as in *Begriffsschrift*. In Frege's terminology, *n*'s being the successor of *m* is expressed as *n*'s "following in the natural series of numbers directly after" *m*.

Prior to providing these analyses, Frege spends a good deal of time in *Grundlagen* arguing for his own view of number-theoretic discourse, in part for the views that numbers are not ideas or other subjectively-determined objects, that they are not properties, and that the role of numerals as singular terms must be taken at face value. The "most fundamental" of his results in this vein, as Frege characterizes it,[31] is the claim that statements of number are assertions about concepts, i.e. that statements of the form "there are *n F*'s" are to be understood as characterizing the concept *F* (as opposed, say, to the collection of *F*'s). The conclusion that attribution of cardinal size to a concept *F* is to be characterized in terms of a relationship borne by a cardinal number to that concept sets the parameters for what Frege will regard as an adequate analysis of the concept *cardinal number*. The central requirement is that the relationship of "belonging to" between a number and a concept be characterized in such a way as to verify what has since become known as "Hume's Principle":

(HP) The number that belongs to *F* = the number that belongs to *G* iff there is a 1–1 map from the *F*'s onto the *G*'s.[32]

Frege's final *Grundlagen* account, on which, again, "the number that belongs to *F*" is taken to stand for the extension of that second-level concept under which fall all and only those first-level concepts equinumerous with *F*, gives exactly the right result: the term "the number that belongs to *F*" ends up as a genuine singular term with respect to which (HP) is straightforwardly verified.

Frege assigns the complex results of his arithmetical analyses, via stipulative definition, as the newly introduced contents of the corresponding ordinary terms (e.g., "zero," "cardinal number"). The familiar terms, as newly understood, are henceforth to be taken strictly as shorthand for Frege's proposed analyses of their ordinary contents. The use of the familiar terminology is of course intended to be suggestive, but can be misleading: it is important to note that the derivations Frege is soon to sketch, in which these terms figure essentially, turn not on the ordinary, pretheoretic contents of the arithmetical terms but on their newly assigned contents, those contents presented by Frege as analyses of the ordinary contents.

Having given the definitions, Frege next sketches the proofs of (what he takes to be) fundamental arithmetical propositions. Section 78, which follows the analyses and definitions, opens as follows:

> I proceed to give here a list of several propositions to be proved by means of our definitions. The reader will easily see for himself in outline how this can be done.
>
> 1. If *a* follows in the natural series of numbers directly after 0, then $a = 1$.
> 2. If 1 is the Number which belongs to a concept, then there exists an object which falls under that concept.
> 3. If 1 is the Number which belongs to a concept *F*; then, if the object *x* falls under the concept *F* and if *y* falls under the concept *F*, *x* is = *y*; that is, *x* is the same as *y*.
> 4. If an object falls under the concept *F*, and if it can be inferred generally from the propositions that *x* falls under the concept *F* and that *y* falls under the concept *F* that *x* is = *y*, then 1 is the Number which belongs to the concept *F*.
> 5. The relation of *m* to *n* which is established by the proposition: "*n* follows in the natural series of numbers directly after *m*" is a one-one relation.
>
> . . .
>
> 6. Every Number except 0 follows in the natural series of numbers directly after a Number.[33]

Sections 79–83 provide a proof-sketch of the crucial principle entailing the infinity of the finite cardinals:

7. After every Number in the natural series of numbers a Number directly follows.[34]

The proof-sketches portray the relevant proofs as relying just on the definitions given and on what Frege takes to be unquestionable laws of logic.[35] The definitions are, as always, dispensable, in the sense that the long-winded versions of (1)–(7) obtained by replacing definienda by definientia throughout are derivable purely from (what Frege takes to be) logical laws.

The analyses and proof-sketches make it likely, says Frege, that the full proofs can be carried out, and hence that it can be demonstrated that arithmetical truths are purely logical. As he puts it:

> I hope I may claim in the present work to have made it probable that the laws of arithmetic are analytic judgments and consequently a priori. Arithmetic thus becomes simply a development of logic, and every proposition of arithmetic a law of logic, albeit a derivative one.[36]

What Frege has in fact shown to be probable (leaving aside the difficulty involving his notion of extension) is that a collection of truths about cardinality-concepts and their extensions are "laws of logic, albeit. . .derivative one[s]." We return in Chapter 4 to the question of whether Frege is right simply to assume, as he does here, that the logical grounding of these truths, i.e. of the output of his careful analyses, suffices to establish the logical grounding of what he primarily cares about, namely, the truths of arithmetic.

1.5 Grundgesetze

By 1893, as sense and reference have replaced the *contents* of Frege's earlier semantic theory, the important questions to raise about the role of analysis are now sharpened. Frege's analytic-cum-definitional strategy is the same here as in *Begriffsschrift* and *Grundlagen*: his central defined terms are explicitly-stipulated abbreviations of complexes of simple terms, which complexes express the results of his analyses of central arithmetical notions. Our term "notions" here is a place-filler: after the instigation of the mature semantic theory, the question arises whether Frege takes his analyses and reductions to be analyses and reductions of the senses or of the references (or neither) of ordinary arithmetical terminology. As we'll see, the answer is not entirely clear.

As in the earlier work, each of Frege's derivations expresses a *proof*, now understood as a series of *thoughts*, where a thought is the sense of a sentence. The final sentence of each important derivation includes, as before, defined terminology, and hence is shorthand for a vastly more complex sentence, one

that expresses the result of a so-called "analysis" of an ordinary arithmetical truth.

The analyses themselves borrow heavily from those given in *Grundlagen*, with some differences. *Grundgesetze*'s scope is larger, including e.g. the beginnings of a theory of real numbers. With respect to the cardinal numbers, the *Grundgesetze* analysis is neater, construing these numbers as the extensions of particular first-level concepts, ones that play, as detailed in Chapter 4, a role very similar to that played by their second-level *Grundlagen* counterparts.

The formal system of *Grundgesetze* is an expanded version of that of *Begriffsschrift*, including, fatally, the notation and axioms for Fregean courses-of-value (*Wertverläufe*), which are generalized versions of *Grundlagen's* extensions. The central purpose of *Grundgesetze* is to give the relevant proofs in full, explicit detail, thereby sealing the case for logicism. As Frege puts it:

> In my *Grundlagen der Arithmetik*, I sought to make it plausible that arithmetic is a branch of logic and need not borrow any ground of proof whatever from either experience or intuition. In the present book this shall now be confirmed, by the derivation of the simplest laws of numbers by logical means alone.[37]

1.6 The General Picture

The general picture painted by Frege of conceptual analysis and its relation to proof is straightforward. Beginning with an ordinary arithmetical truth of the kind typically expressed via a familiar arithmetical sentence, Frege subjects that truth to careful analysis, usually by breaking down its important components into what he takes to be simpler ones. The new, highly analyzed version of the original truth is expressed either by a sentence whose syntactic complexity is greater than that of the familiar sentence, or by a definitional abbreviation thereof. Frege's strategy is then to derive that sentence using only logic, which is to say that he proves the highly analyzed version of the arithmetical truth from purely logical principles. Finally, Frege's central claim is that the proofs of his highly analyzed versions of arithmetical truths demonstrate the purely logical grounding of arithmetic, and thereby substantiate his thesis that arithmetic is a branch of logic.[38]

As noted above, Frege's analyses prior to *Grundlagen* deliver contents that are relatively easily recognizable clarifications of their ordinary analysanda. In the early cases like those canvassed in §§1.22 and 1.3, a proof of the analysans-content from premises provides, arguably, all one could ask for as a demonstration that the analysandum follows logically from those premises. By the time of *Grundlagen*, however, the analyses are considerably less straightforward: it is no longer the case, by 1884, that the contents and thoughts Frege proves are easily

recognizable clarifications of the ordinary arithmetical truths they are taken in some sense to represent. Frege's *Grundlagen* version of

(β) One is the successor of zero

is roughly

(β*) For some concept *F* and some object *a* falling under *F*, the extension of the concept *equinumerous with the concept identical with the extension of the concept (equinumerous with the concept (not self-identical))* is the extension of the concept *equinumerous with the concept F*, and the extension of the concept *equinumerous with the concept not self-identical* is the extension of the concept *equinumerous with the concept falling under F, but not identical with a*.

The immediate question is that of how Frege understands the similarity between analysandum and analysans in general. If we take him at his word, Frege holds that a proof of the content of (β*) from purely-logical principles demonstrates that the content of (β) is grounded in just such principles. Similarly for the rest of what he takes to be the fundamental truths of arithmetic. But given the considerable distance (as explored in the next chapter) in each case between the ordinary arithmetical truth and its highly-analyzed, proven counterpart, it is a significant question what he thought his analyses were providing, and in virtue of what kind of similarity-relation between analysandum and analysans he took the latter to function as a reliable stand-in for the former.

Given the importance of analytic accuracy within his project, one might expect Frege to provide some clear guidelines concerning analysis in general. Notoriously, however, he offers nothing of the sort. Though he repeatedly affirms the central importance of analysis, Frege never defines this notion, and—most importantly—never gives an account of the necessary and sufficient conditions for successful analysis. We are left to look at the details of the analyses he gives, and at his various scattered remarks about the nature of analysis, to determine what he took himself to be providing under the heading of "analysis" or "elucidation" and to assess the plausibility of his general approach.

Despite Frege's lack of clarity about the nature of analysis, one thing about which we can be clear at this point is as follows. Frege's failure to lay down criteria of successful analysis, and the consequent imprecision regarding the boundaries of the relation of successful analysis, do not impugn the rigor of Fregean proof. No step in a proof is ever justified by appeal to the correctness of a conceptual analysis. Appeal to such correctness takes place when one takes the proof of an analysans-thought from premises *P* to demonstrate that the corresponding analysandum-thought is itself provable from, or grounded in, *P*. In the case of most

concern to Frege, the crucial appeal to analytic correctness takes place when one takes the rigorous proofs in *Grundgesetze*, proofs of highly-analyzed versions of arithmetical truths, to demonstrate the logical grounding of the arithmetical truths themselves. But again, the appeal to analytic correctness is relevant not to the rigor of Frege's proofs, but to their significance.

To get a clear idea of just how Frege understands the kinds of conceptual analysis involved in his work, and the relationship between the thoughts he proves and the thoughts expressed by ordinary arithmetical discourse, it will be important to lay out clearly some of Frege's central views about thoughts and about their expression in both natural and formal languages. This is where we turn next.

Notes

1. *Begriffsschrift* (hereafter *Bff*) Preface pp iii–iv/5.
2. The system of *Begriffsschrift* doesn't yet have the problematic involvement with extensions or courses-of-value that later dooms Frege's formalism; it is consistent, and its principles are arguably purely logical. Only those who object to higher-order quantification will take the bounds of the purely logical to have been breached at this point. The system becomes inconsistent in 1893 with the addition of terminology for, and axioms governing, courses-of-value.
3. *Bff* Preface p viii/8.
4. *Grundgesetze* (hereafter *Gg*) vol. 1 §4. On Frege's use here of the word "definition," see Section 1.4.
5. Ibid., emphasis added.
6. *Grundlagen* (hereafter *Gl*) §46.
7. *Gl* §90.
8. See *Gg* 1 Introduction pp viii–ix/5:

   > With this book I carry out a design that I had in view as early as my *Begriffsschrift* of 1879 and announced in my *Grundlagen der Arithmetik* of 1884. I wish here to substantiate in actual practice the view of Number that I expounded in the latter book. The most fundamental of my results I expressed there, in §46, by saying that a statement of number expresses an assertion about a concept; and the present account rests upon this.

9. [1914] p 226/209.
10. See Weiner [1990], [2007].
11. As van Heijenoort points out, Frege also uses an unstated rule of substitution. See van Heijenoort's introductory notes to *Bff* in van Heijenoort [1967b] p 2.
12. See [1882a] pp 97–98/90–91:

    > I did not wish to present an abstract logic in formulas, but to express a content through written symbols in a more precise and perspicuous way than is possible with words.

    Also *Bff* Preface:

    > I have omitted the expression of everything which is without importance for the chain of inference. In §3, I have designated by *conceptual content* (*begrifflicher*

> *Inhalt*) that which is of sole importance for me. Hence, this must always be kept in mind if one wishes to grasp correctly the nature of my formula language.

13. *Bff* Preface p iv/6.
14. *Bff* §24. This view of definitions remains constant throughout Frege's work. We see below its occurrence in *Grundgesetze*, Frege's next formal work. Later, in the posthumously published "Logic in Mathematics" written in 1914, we find: "Definitions are not absolutely essential to a system. We could make do with the original group of signs. The introduction of a simple sign adds nothing to the content; it only makes for ease and simplicity of expression" (p 224/208).
15. In *Grundgesetze*, the same point is put as follows: "The definitions . . . merely introduce abbreviated notations (names), which could be dispensed with were it not that lengthiness would then make for insuperable external difficulties" (*Gg* 1 Introduction, p vi/2).
16. *Bff* §23.
17. *Bff* §24. The triple-bar is Frege's own sign, which he takes in 1879 to stand for the relation of content-identity, a relation between signs. Frege's right-hand side is syntactically more complex than is ours, in order that the same terminology can be used in a slightly different arrangement to express what we would have to render by using new terminology for the inverse of the relation *f*.
18. Here "predicate" is understood broadly to include the result of removing the appropriate number of singular terms from any well-formed formula.
19. Again, Frege's right-hand side is syntactically more complex than ours; we have sacrificed faithfulness to his more versatile presentation for readability.
20. An *f*-successor of *x* is any *y* such that *f(x,y)*.
21. *Bff* §28. The label "(TRANS)" is added.
22. *Bff* §27 note.
23. *Bff* Preface p iv/5.
24. Unusual, but not unheard of. Poincaré argues against Couturat's similar account that the equivalence between (roughly) (α) and (α^*) relies on an understood principle of induction and hence that the claim of equivalence is circular. See Poincaré [1905]; for discussion see Goldfarb [1988]; for a reply to the objection, see Demopoulos [1994].
25. George Boolos [1985] argues that Frege's derivation of (98) is not the strongest example in *Begriffsschrift*'s Part III of this anti-Kantian claim. As Boolos sees it, the Kantian can reply that while the content of (98) is indeed grounded in pure logic, it's not one of those that appears *prima facie* to require appeal to intuition for its verification. The decision as to the justice of the Kantian claim will turn, as Boolos points out, on the adequacy of the analysis of (TRANS) in terms of (98). Boolos notes here that a considerably stronger anti-Kantian case is made by Frege's derivation of (133), which says that if *f* is a function, then any two objects *a* and *b* bearing the ancestral of *f* to a given object *x* will be connected by the ancestral of *f*.
26. *Bff* Preface p viii/8.
27. [1881] pp 30–31/27–28. (P1) and (P2) are formulas (1) and (2), respectively, of [1881]. (SUM) is Frege's own natural-language rendering of the claim to be proven; see [1881] p 30/27. Addition and zero are left unanalyzed in (P1) and (P2).
28. This is Frege's (12) in modern notation. See [1881] p 34/31.
29. *Gl* §4.
30. See [1914] pp 226–227/209–210.
31. *Gg* 1, Introduction pp viii–ix/5.
32. See *Gl* §73. The name is due to George Boolos and alludes to Frege's *Gl* §63 citation of Hume as having held something like the principle. See Boolos [1990].
33. *Gl* §78.
34. The labels 1–6 are Frege's. The final proposition quoted, from §79, is not numbered by Frege; we have added the label 7. The infinity of the collection of finite cardinals follows because it is also provable via Frege's definitions that (as above) the ancestral of the

successor-relation is transitive and that it is irreflexive on the collection composed of 0 together with those numbers bearing the ancestral of the successor relation to 0.

35. These laws include laws of ordinary first- and second-order logic, together with the problematic (and not explicitly mentioned) existence and comprehension principles for extensions. See Boolos and Heck [1998] for problems with Frege's proof-sketch of (7).
36. *Gl* §87. See also *Gl* §109: "From all the preceding it thus emerged as a very probable conclusion that the truths of arithmetic are analytic and a priori; and we achieved an improvement on the view of Kant." As Frege uses the term in *Grundlagen*, "analytic" truths are those provable from logical laws and definitions (see *Gl* §3), which, given the nature of definitions, establishes that these truths are, as Frege says, themselves truths of logic. Frege later drops his reliance on the category of analyticity, claiming straightforwardly that the truths of arithmetic are provable from logical principles and hence are themselves purely logical.
37. *Gg* 1 §0.
38. See *Gl* §87.

2

Thoughts

2.1 Thoughts and Language

Frege insists throughout his career on the importance of distinguishing linguistic from nonlinguistic items, and is particularly concerned with the nonlinguistic entities expressed by complete sentences. In *Begriffsschrift*, as we've seen, the things expressed by sentences are called variously "conceptual contents" or "contents of possible judgment"; the latter phrase is sometimes abbreviated simply as "judgment" or "content." Later, by the time of *Grundgesetze*, the nonlinguistic entities expressed by sentences have become considerably more-carefully explained, and have come to play a central role in Frege's general semantic theory. Now known as *thoughts* (*Gedanken*), these entities are the centerpiece of Frege's famous two-tiered theory of sense and reference, serving as the senses of complete declarative sentences.

Despite a number of changes between the early and later views, the central role with respect to the logicist project of the entities expressed by sentences is unchanged. The *conceptual contents* of *Begriffsschrift* and the *thoughts* of *Grundgesetze* are what mathematical, and other, theories are made of. They are the objects of knowledge and belief, the bearers of truth-value, and the things between which the relation of logical entailment obtains. An important reason for taking such things seriously, from Frege's point of view, is that they offer a clean account of common intellectual content: when we say that A believes or doubts the *same thing* as does B, what we say is on Frege's account literally true: there is a single thought to which both A and B are appropriately related. "[O]ne can hardly deny," says Frege, "that mankind has a common store of thoughts which is transmitted from one generation to another."[1] Similarly,

> Can the same thought be expressed in different languages? Without a doubt, so far as the logical kernel is concerned; for otherwise it would not be possible for human beings to share a common intellectual life. But if we think of the kernel with the psychological husk added, a precise translation is impossible. Indeed we may go so far as to doubt whether the outer covering is the same for any two men. From this we can see the

> value of learning foreign languages for one's logical education; when we see that the same thought can be worded in different ways, our mind separates off the husk from the kernel, though, in any given language, it appears as a natural and integral part of it. This is how the differences between languages can facilitate our grasp of what is logical.[2]

> Even if with us men thinking does not take place without ideas, still the content of a judgment is something objective, the same for everybody, and as far as it is concerned it is neither here nor there what ideas men have when they grasp it.[3]

Arithmetical truths are, as Frege sees it, such nonlinguistic entities. Conceptual contents or thoughts are the things one takes a proof to have proven, and hence they are the things whose purely-logical grounding, in the case of arithmetic, Frege hopes to demonstrate:

> When one uses the phrase "prove a proposition" in mathematics, then by the word "proposition" we clearly mean not a sequence of words or a group of signs, but a thought; something of which one can say that it is true.[4]

The question raised at the end of the last chapter—that of the role in Frege's logicist project of his conceptual analyses—can be put in terms of conceptual contents or thoughts. The question is this: what is the relationship between the conceptual contents or thoughts Frege actually proves, i.e. those delivered as the end result of his often highly nontrivial analyses, and the conceptual contents or thoughts of ordinary arithmetic, those expressed by such sentences as "Every cardinal number has a successor," and that serve as the starting-point of his analyses? To answer this question is to answer the question of what Frege took an adequate conceptual analysis to be, at least in this central case.

We need, to begin with, a clear account of the nonlinguistic entities in question and of the role they play in Frege's conceptions of logic, of mathematics, and of scientific reduction. We'll focus on the more well-worked-out version of Frege's picture, and hence on the mature *thoughts* of the post-1890 period. But it will be important to see how thoughts emerge from the earlier conceptual contents, and to be clear about the role played by these contents in *Begriffsschrift* and *Grundlagen*, before turning to the mature theory.

In *Begriffsschrift*, Frege takes it that the things he is interested in, the truths of arithmetic, are the kinds of things one can affirm or deny. That is, they are the kinds of things that can form the content of *judgments*. His picture of the relationship between a formula of *Begriffsschrift* and such a judgment-content is as follows: a formula itself "indicates" or is a "sign" for a judgment-content; a formula preceded by his special assertion-sign is to be taken to express the judgment

whose content is indicated.[5] Judgment-contents themselves are rather vaguely construed sometimes as "combinations of ideas," ideas that stand as the *contents* of individual signs or other sentence-parts, and sometimes as constituted in part by objects (e.g. geometrical points), which objects stand as the contents of individual signs.[6] The kinds of contents that can form the content of a judgment are always clearly distinguished from the non-judgeable contents of the kind indicated by sub-sentential signs.

> Not every content becomes a judgment when [the assertion-sign] is written before its sign; for example, the idea "house" does not. We therefore distinguish *judgeable* from *non-judgeable* contents. [footnote: On the other hand, the circumstance that there are houses, or that there is a house . . . is a judgeable content. But the idea "house" is only a part of it. . . .][7]

The 1879 discussions of judgeable contents are relatively rudimentary and, as just noted, not entirely consistent with respect to the composition of contents. Frege clearly feels no need to lay out in any detail a theory of these entities. The important features of his view at this point, and those that carry over into the mature semantic theory, are (1) that the purpose of sentences, whether of German or of his formal language, is simply the expression of the kinds of things that form the contents of judgment; and (2) that these contents, and not the sentences themselves, are the things we're interested in when we ask about the "grounds" of a science or about relationships of logical entailment or of provability.[8]

Frege gives us in the early years of his project no clear criteria of identity for judgeable contents. That is, he gives no clear conditions under which two sentences should be said to express the same such content. The closest he comes is the following passage, again from *Begriffsschrift*:

> A distinction between *subject* and *predicate* does *not occur* in my way of representing a judgment. In order to justify this I remark that the contents of two judgments may differ in two ways: either the consequences derivable from the first, when it is combined with certain other judgments, always follow also from the second, when it is combined with these same judgments, or this is not the case. The two sentences "The Greeks defeated the Persians at Plataea" and "The Persians were defeated by the Greeks at Plataea" differ in the first way. Even if one can detect a slight difference in meaning, the agreement outweighs it. Now I call that part of the content that is the *same* in both the *conceptual content*. Since *it alone* is of significance for our concept-script, we need not introduce any distinction between propositions having the same conceptual content. . . . Now, all those peculiarities of ordinary language

> that result only from the interaction of speaker and listener—as when, for example, the speaker takes the expectations of the listener into account . . . have nothing that answers to them in my formula language, since in a judgment I consider only that which influences its *possible consequences*.[9]

The central idea here seems to be that, from the array of semantic features had by sentences, we can separate a core, the *conceptual content*, a core that includes everything relevant to logical entailment. Frege's view that the conceptual content includes *only* what's relevant to logical entailment arguably implies that in his view all logically-equivalent sentences express the same conceptual content.[10] As we'll see, this idea crops up at various times throughout Frege's work, but his degree of commitment to the view would seem to vary from context to context.

There is little development of the semantic theory over the next five years, in the period in which *Grundlagen* is written. In *Grundlagen*, Frege again talks in a relatively-unfocused way about "contents" and seems to take it that the truths (or falsehoods) expressed by arithmetical sentences are composed of objects, concepts, and relations more generally.

By the time of the first volume of *Grundgesetze* in 1893, Frege has developed his two-tiered semantic theory, on which every piece of language (formal or natural) has a sense and (typically) a reference. The mature theory of sense is most clearly articulated in the well-known trio of papers published in 1891 and 1892.[11] Here Frege often introduces senses in connection with what he calls, non-technically and intuitively, the "information" expressed by a sentence. In discussing his own claim that the sentence "The horse is a four-legged animal" in its typical use expresses just what "all horses are four-legged animals" does, Frege remarks:[12]

> Nowadays people seem inclined to exaggerate the scope of the statement that different linguistic expressions are never completely equivalent, that a word can never be exactly translated into another language. One might perhaps go even further, and say that the same word is never taken in quite the same way even by men who share a language. I will not enquire as to the measure of truth in these statements; I would only emphasize that nevertheless different expressions quite often have something in common, which I call the sense, or, in the special case of sentences, the thought. In other words, we must not fail to recognize that the same sense, the same thought, may be variously expressed; thus the difference does not here [i.e. with respect to the example just noted—PB] concern the sense, but only the apprehension, shading, or coloring of the thought, and is irrelevant for logic. It is possible for one

> sentence to give no more and no less information than another; and, for all the multiplicity of languages, mankind has a common stock of thoughts. If all transformation of the expression were forbidden on the plea that this would alter the content as well, logic would simply be crippled; for the task of logic can hardly be performed without trying to recognize the thought in its manifold guises. Moreover, all definitions would then have to be rejected as false.[13]

One noteworthy feature of this introductory explanation is Frege's idea that thoughts are in some sense what's *common*, under the right circumstances, to the relatively rich and fine-grained complex of semantic features expressed by each of a collection of sentences. The idea is that despite differences in e.g. the poetic quality or "coloring" of what's expressed by the sentences in question, there is nevertheless an overlapping core of similarity, and that this core (i.e. what's left when we disregard the extraneous semantic features) is the thought expressed. As we'll see, this conception of the thought as a common element in, and one extracted as it were from the richer total semantic correlate of a collection of semantically-similar sentences, seems to remain constant throughout Frege's work, though his account of which parts of that semantic correlate are "extraneous" to the thought expressed is arguably less constant.

The second feature worth noting from this passage is Frege's appeal, as above, to the intuitive idea that one sentence can on occasion "give no more and no less information than another," and that this is the circumstance under which the two will be said to express the same thought.[14] The notion of information itself is left relatively unclear at this point, but is repeatedly cashed out by Frege in terms both of what's "relevant for logic," as in the passage just quoted, and of what's relevant with respect to assent and dissent. With respect to the latter, Frege provides some clarification in this period via his often-cited discussions of speakers' attitudes toward sentences. Frege tells us, for example:

> If we say "the Evening Star is a planet with a shorter period of revolution than the Earth," the thought we express is other than in the sentence "the Morning Star is a planet with a shorter period of revolution than the Earth;" for somebody who does not know that the Morning Star is the Evening Star might regard one as true and the other false.[15]

Presumably what's meant here is that, restricting ourselves to people who understand both sentences, a person might reasonably or coherently do as Frege says, i.e. take the sentences to have different truth-values. Though the condition Frege implicitly appeals to here is only a necessary and not a sufficient condition for thought-identity (and a vague one at that), this discussion and the surrounding

examples hint at what one might call the "cognitive" criterion of thought-identity, namely, that two sentences express the same thought iff a speaker who understands both of them and assents to one must, on pain of incoherence, also be disposed to assent to the other. As Frege puts it in an unpublished manuscript of ca. 1906:

> [T]wo sentences A and B can stand in such a relation that anyone who recognizes the content of A as true must thereby also recognize the content of B as true and, conversely, that anyone who accepts the content of B must straightaway accept that of A. (*Equipollence*). It is here being assumed that there is no difficulty in grasping the content of A and B. . . . I assume there is nothing in the content of either of the two equipollent sentences A and B that would have to be immediately accepted as true by anyone who had grasped it properly. . . .
>
> So one has to separate off from the content of a sentence the part that alone can be accepted as true or rejected as false. I call this part the thought expressed by the sentence. It is the same in equipollent sentences. . . .[16]

One of the difficulties in applying the cognitive criterion to a pair of sentences in order to determine whether they express the same sense arises in those cases, for which Frege explicitly makes allowances here, in which the sentences in question express self-evident truths. The cognitive criterion in full generality would imply that all such sentences express the same thought. Frege's concession to such cases in this passage is simply that they are exceptions to the cognitive criterion: though pairs of such sentences are in fact such that anyone understanding both and assenting to one must (on pain of incoherence) assent to the other, they are not to be taken for that reason to express the same thought. Taking the criterion given here as our only guide to sense-identity, then, we have no account of the conditions under which sentences expressing self-evident truths express the same thought. This point will become particularly important below, when we turn to arithmetic, in virtue of what is arguably the self-evidence of some logical and arithmetical truths.

A perhaps more-interesting difficulty concerns the conditions under which a speaker is said to have "grasped . . . properly" the content of a sentence. If a speaker is reasonably imagined to assent to one sentence while coherently dissenting from or claiming agnosticism about another, the question of whether this demonstrates that the two sentences express different thoughts will turn on the question of whether the imagined speaker can also reasonably be said to have understood those sentences, that is, to have grasped their contents properly. Some cases are easy: the speaker who assents to "The Americans destroyed a sovereign nation" while dissenting from "A sovereign nation was destroyed by

the Americans" evidences a clear lack of understanding; the cognitive criterion tells us that the sentences express the same thought. But what about a speaker who assents to "5 + 5 = 10" and dissents from "2 times 5 = 10"? Does such a speaker give evidence of a distinction between the thoughts expressed, or of a failure to grasp the sense of the phrase "2 times . . ."? To the extent that this question has no determinate answer, the cognitive criterion will give no guidance as to the identity of the thought(s) expressed by such pairs of sentences.

The cognitive criterion, such as it is, is put under a good deal of pressure when conjoined with Frege's view of thoughts as forming the content of science, and of underwriting the continuity of specific sciences and theories over time, as being passed on, as Frege puts it, "from one generation to another." In virtue of what, one might want to know, is it the case (if it is the case) that e.g. Newton and Weierstrass were engaged in a common science? What, in general, makes it the case that two researchers across vast gulfs of language, time, and theoretical presuppositions can nonetheless engage the same mathematical or other scientific questions? Perhaps the right thing to say here is just that the two researchers in each case stand in an appropriate, if hard-to-identify, historical connection to one another. From the Fregean point of view, this historical connection will be spelled out in terms of the thoughts entertained, believed, proven, and so forth by each researcher, and the connections between them across time. Given Frege's remarks to the effect that our possession of "common sciences" turns on our possession of a "common stock of thoughts," a natural Fregean approach is to say that the central thoughts in question are identical: Newton and Weierstrass apprehended the same thoughts even in cases in which the manner of their understanding was significantly different.

This first strategy—that of holding that e.g. Newton and Weierstrass entertained some of the same significant mathematical thoughts—gives the smoothest account along Fregean lines of the idea that two such researchers can agree or disagree with one another: they are simply affirming, on this account, the same thoughts or negations thereof. No other account, if one has already plumped for a picture of judgment in terms of the affirmation of something like thoughts, can have the result that mathematicians of significantly-different eras can literally agree or disagree with one another.[17] And as noted above, the picture of temporally-distant scientists investigating the same thoughts fits well with various of Frege's remarks regarding the possibility of a "common science," and of his repeated claims that the theorems of a science are a determinate collection of thoughts.[18] The difficulty with this approach from the Fregean point of view is that it doesn't sit neatly with the fine-grained individuation of thoughts given by the cognitive criterion. For on that criterion, Newton's and Weierstrass's sentences express the same thoughts only if they are something like easily-recognizable synonyms, sentences with respect to which it would be obviously incoherent to affirm one without affirming the other. And this strong kind

of semantic equivalence is, it seems, considerably too demanding: whatever mathematicians aim for over the course of mathematical development, obvious synonymy with their predecessors' sentences is clearly not it. There is more to be said here, particularly as this issue arises with respect to Frege's own proposed development of arithmetic. For now, the point is just this: there is on the face of it at least a certain tension between Frege's best-known criterion of thought-identity and his view of theories as collections of thoughts that survive significant conceptual development.

There is occasional evidence in Frege's texts of a different criterion of thought-identity, one more closely tied than is the cognitive criterion to the role of thoughts as the subject-matter of logic. This criterion appears first in the early theory of judgment-contents. In *Begriffsschrift*, as we've seen, judgment-contents are taken to include just what's relevant to logical entailment; in the same work, Frege notes as well that in expressing conceptual contents his formulas express just what is of "significance for the *inferential sequence*."[19] Similarly, but much later, Frege remarks in a 1906 letter to Husserl:

> It seems to me that an objective criterion is necessary for recognizing a thought again as the same, for without it logical analysis is impossible. Now it seems to me that the only possible means of deciding whether sentence A expresses the same thought as sentence B is the following, and here I assume that neither of the two sentences contains a logically self-evident component part in its sense. If *both* the assumption that the content of A is false and that of B true *and* the assumption that the content of A is true and that of B false lead to a logical contradiction, and if this can be established without knowing whether the content of A or B is true or false, and without requiring other than purely logical laws for this purpose, then nothing can belong to the content of A as far as it is capable of being judged true or false, which does not also belong to the content of B; . . . In the same way, given our supposition, nothing can belong to the content of B, as far as it is capable of being judged true or false, except what also belongs to the content of A. Thus what is capable of being judged true or false in the contents of A and B is identical, and this alone is of concern to logic, and this is what I call the thought expressed by both A and B. One can indeed count many sorts of things as part of the content of A, e.g. a mood, feelings, ideas; but none of these is judged true or false; at bottom it is of no concern to logic, just as whatever is incapable of being judged morally good or bad is of no concern to ethics.[20]

We'll have more to say below about the opening remark concerning the connection between "logical analysis" and thought-identity. For present purposes, the

interesting point of this passage is that here the connection to logic seems to eclipse the connection to judgment in the account of the conditions of thought-identity. Leaving aside any sentence that "contains a logically self-evident part in its sense," whatever exactly this means, Frege's idea seems to be that if two sentences express logical equivalents, then they express the same thought.

There are several difficulties with this criterion, primarily a certain potential circularity together with the difficulty that it would seem to contradict a number of Frege's own examples of sentences that he takes to express different thoughts.[21] Given his logicism, true arithmetical sentences express logical equivalents, but Frege explicitly holds of various such sentences that they express distinct thoughts.[22] This last difficulty may in fact be part of what Frege is trying to avoid in his stipulation that the criterion isn't to be applied to sentences involving "self-evident" parts. But if so, then, just as with the cognitive criterion, this condition of thought-identity simply tells us nothing about the kinds of cases we'll be most interested in, namely, those having to do with fundamental logical and arithmetical truths and falsehoods.

To sum up the discussion to this point: Frege holds throughout his career that the central role of sentences, whether of formal or of natural languages, is to express the kinds of things that can form the contents of judgment. He seems to hold in the early works, and certainly holds in the later, a compositional account of these nonlinguistic entities: the content, or the sense, of a complex linguistic item is determined by the contents, or the senses, of its constituents. With respect to the well-developed later theory: the entities expressed by, i.e. the senses of, complete declarative sentences are also known as *thoughts*. A given thought can be expressed by different sentences; indeed, the fact that a thought can be expressed by different sentences is what explains the possibility of a "common science" across linguistic divisions. In both the earlier and the later parts of his career, Frege holds that the entities in question—that is, the earlier conceptual contents or the later thoughts—are the things between which the relation of logical entailment, and the associated relations of consistency, inconsistency, and so on, obtain. Similarly for the relation of provability.

With respect to the mature theory, there are clear cases of sentence-pairs that express on Frege's view the same thought. Here, for example, a pair of sentences differing just with respect to active versus passive grammatical construction will certainly count. There are also clear cases of sentence-pairs that express on Frege's view different thoughts. Any sentences with different truth-values will of course express different thoughts; those with evidently different truth-conditions will presumably do so as well. As we have seen, there are also sentence-pairs between which the Fregean criteria we have examined so far give no clear answer to the question of thought-identity.

We return to the issue of thought-identity in Chapter 4. But first we turn here to a handful of related questions regarding sense and reference.

2.2 Sense and Reference

According to Frege's mature semantic theory, each significant linguistic item has not just a sense but typically also a reference. The references of singular terms include both ordinary objects like desks and mountains, and abstract objects like numbers. Predicative phrases, which category includes any significant portion of a sentence obtained by removing one or more terms from that sentence, refer to what Frege calls "functions." The peculiarity of a function is that, unlike an object, it is essentially predicative: to refer to a function is to predicate. The fact that a sentence "hangs together" to make a claim, instead of standing simply as a list of names, is due on Frege's account to the fact that its predicative phrases refer to these predicative things, i.e., to functions. In a metaphor that appears throughout his work, Frege refers to singular terms and their references as "saturated" or "capable of standing alone," and to predicative phrases and their references as "unsaturated," "in need of supplementation."[23] The sentence "7 is prime" differs semantically from e.g. the list of terms <"7," "primeness">, or <"7", "possession," "primeness"> because as Frege sees it the phrase ". . . is prime" refers to a function, the kind of thing reference to which amounts to predication rather than to naming.[24]

The reference of a complex phrase is determined functionally by the references of its parts. "The successor of 7" refers to the value of the function referred to by "The successor of . . ." as applied to the reference of "7" as argument. "The father of the father of Gottlob Frege" refers to the value of the function referred to by "the father of . . ." on the argument referred to by the term "the father of Gottlob Frege," which in turn is the value of the function referred to by "the father of . . ." on the argument referred to by the term "Gottlob Frege." The functions referred to here are just what one would expect: the function in our first example is one that delivers numbers for numbers as arguments, and in the second, one that delivers humans for humans as arguments. Frege takes it that one-place predicates whose argument-places are appropriate to be filled with singular terms refer to functions whose arguments are of course objects, and whose values are truth-values. As one might expect, for example, the function referred to by ". . . is prime" takes one from prime numbers to the value *true* and from other objects to *false*.[25] In keeping with the general thesis that the reference of a complex linguistic item is the value of the function(s) referred to by its function-parts as applied to the arguments referred to by its argument-parts, Frege takes it that the reference of a declarative sentence is its truth-value.

That a simple piece of language has a sense as well as a reference is indicated, as Frege sees it, by the fact that the role of that piece of language in conveying information goes beyond just that of referring. We can see this by noting that two terms with the same reference can nonetheless play different roles in the

expression of that kind of thing that forms the content of a judgment or assertion. Frege's "morning star"/"evening star" example is the paradigm of such a pair of terms.[26] Though as we'll see soon Frege is not entirely clear on this point, he often writes as if two co-referring terms will have different senses whenever it takes a nontrivial amount of investigation (which can be entirely *a priori*) to determine that they do co-refer. In keeping with this way of thinking of the issue, Frege sometimes writes of the sense of a singular term as containing a "mode of presentation" of its reference: the senses of "the morning star" and of "the evening star" contain two distinct modes of presentation of the planet Venus.[27]

The reference of a simple piece of language is determined by its sense, at least in the minimal sense that if two simple pieces of language have the same sense, then they have the same reference. This, together with the fact that the sense of a complex linguistic item is determined as above by the senses of its parts and that the reference of a whole is determined as noted earlier by the references of its parts, implies the general Fregean thesis that the reference of any piece of language is determined by its sense: if two linguistic items have the same sense, they have the same reference. Consider again the sentence "7 is prime." Its sense is the thought obtained by putting together the sense of "7" and the sense of ". . . is prime"; this thought is true, which means that the sentence itself refers to the truth-value *true*. Equivalently, the reference of ". . . is prime," a function determined by the sense of that predicate, when applied to the reference of "7," an object determined by the sense of "7," delivers as value the truth-value *true*, which has again the result that the sentence as a whole refers to that truth-value.

Inspired in part by Frege's occasional talk of senses as containing "modes of presentation" of references, it is sometimes thought that on the Fregean view a linguistic item can have sense only if it has reference.[28] But this is a mistake. It is essential to Frege's view, and a point about which he is clear, that some pieces of language have sense despite lacking reference. As Frege says:

> It may perhaps be granted that every grammatically well-formed expression figuring as a proper name always has a sense. But this is not to say that to the sense there also corresponds a reference. The words "the celestial body most distant from the Earth" have a sense, but it is very doubtful if there is also a thing they refer to. The expression "the least rapidly convergent series" has a sense but demonstrably there is nothing it refers to, since for every given convergent series, another convergent, but less rapidly convergent, series can be found. In grasping a sense, one is not certainly assured of referring to anything.[29]

Fiction too provides as Frege sees it cases of terms that have sense without reference. The term "Odysseus," for example, assuming that the *Odyssey* is a work

of fiction, has a sense but no reference, as do sentences in which that term appears in non-oblique position.[30]

Frege does claim that, in the case of artificially-designed formal languages, ones intended for the rigorous presentation of proofs, it is important that the kinds of case just mentioned never occur.[31] In a good formal language, no well-formed terms lack reference, and no well-formed sentences lack truth-value. Hence no good formal language will contain a term with the sense of the English-language terms "the least rapidly convergent series" or "Odysseus." We turn in detail in Chapter 3 to the reasons for this requirement on formal languages, but the simple outline is just that if a sentence of a formal language has no truth-value (as will generally be implied by the existence in that language of a term lacking reference), then it is impossible to give a syntactic specification of a strictly truth-preserving deductive system for that language. In the context of a Fregean formal language, possession of sense by a well-formed piece of language suffices for its possession of reference, for the simple reason that its well-formedness alone suffices, by design, for its possession of reference. But this again is a feature had by good formal languages as a matter of design, and is not in Frege's view a feature of intelligible language in general.

2.3 The Structure of Sense

As noted, Frege holds both that the thought expressed by a sentence is determined by the senses of its parts and their order of composition, and that distinct sentences can express the same thought. With respect to these minimal theses, his view is just that of virtually any propositions-theorist. But Frege holds in addition the less-common view that sentences of radically different syntactic structure can yet express the same thought. This view has its roots in Frege's early semantic theory. In *Grundlagen*, Frege tells us that the content expressed by a sentence can be "carved up" in different ways, and that this "recarving" can yield "new concepts." What he means by this is best illustrated by his own example.[32] The sentence

(par) Line *a* is parallel to line *b*

expresses, as Frege sees it, the same content as does

(dir) The direction of line *a* = the direction of line *b*.

Taking their superficial syntactic structure as a guide to the contents expressed, one would be led to take (par) as expressing a content composed

out of, or otherwise essentially linked to, two lines and the relation of parallelness, and (dir) as expressing a content similarly tied to two quite different objects, *directions*, and the relation of identity.[33] One might further be inclined on this basis to suppose either (1) that the sentences express different contents, or (2) that the superficial grammatical structure of at least one of them is, after all, misleading. Frege's view is the perhaps-startling one that the superficial syntactic structure is in both cases a reliable guide to the composition of the content expressed and that the two sentences express the same content. The content is just, on this view, decomposable in a variety of ways. As Frege puts it:

> The judgment "line *a* is parallel to line *b*," or, using symbols,
>
> $a//b$
>
> can be taken as an identity. If we do this, we obtain the concept of direction, and say: "the direction of line *a* is identical with the direction of line *b*." Thus we replace the symbol // by the more generic symbol = , through removing what is specific in the content of the former and dividing it between *a* and *b*. We carve up the content in a way different from the original way, and this yields us a new concept. Often, of course, we conceive of the matter the other way round, and many authorities define parallel lines as lines whose directions are identical. . . . Only the trouble is, that this is to reverse the true order of things. For surely everything geometrical must be given originally in intuition. But now I ask whether anyone has an intuition of the direction of a straight line. Of a straight line, certainly, but do we distinguish in our intuition between this straight line and something else, its direction? That is hardly plausible. The concept of direction is only discovered at all as a result of a process of intellectual activity which takes its start from the intuition.[34]

In short, we first acquire an understanding of the content of sentences of the form (par), together with an understanding, given via intuition, of the contents of their singular terms and of their relation-term ". . . is parallel to. . . ." We subsequently arrive at an understanding of the more abstract terms of the form "the direction of. . . ." via a recognition of the fact that sentences of the form (dir) express the same content as do their counterparts of the form (par). Directions are recognized, in short, as what parallel lines have in common, and the Fregean explanation of this, at least in 1884, is that direction-identity statements share content with parallel-line statements.

This view that contents are multiply decomposable forms an important part of Frege's semantic theory, appearing as early as *Begriffsschrift* and surviving the change to the mature version of that theory. At *Begriffsschrift* §9, Frege writes:

> "The circumstance that carbon dioxide is heavier than hydrogen" and "The circumstance that carbon dioxide is heavier than oxygen" are the same function with different arguments if we regard "hydrogen" and "oxygen" as arguments; on the other hand, they are different functions of the same argument if we regard "carbon dioxide" as the argument.
>
> . . .
>
> [Consider] the proposition that Cato killed Cato. 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.

The picture one gets here is that of a content as put together like a jigsaw puzzle, decomposable in various ways into clusters of smaller pieces. On this view, the fact that two sentences indicate quite different ways of decomposing their whole content into parts is no reason to suppose that the contents expressed by those sentences are distinct. The re-carvings noted in these *Begriffsschrift* examples are not as radical as is the re-carving that gets us from (par) to (dir): in the *Begriffsschrift* examples, each object and function can be seen as the reference of an appropriate (if gappy) cluster of terms in the original sentence, while no cluster of terms in the sentence (par) stands for the direction of line *a*. But implicit in the final *Begriffsschrift* remark quoted is the idea that the content of "Cato killed Cato" is the same as that of "Cato committed suicide." The move from the decomposition indicated by the second of these sentences to that indicated by the first is again one in which the newly-discovered concept (e.g. *to kill Cato*) is not the reference of any cluster of terms appearing in the original.

After the advent of the mature semantic theory, the entities multiply decomposable are thoughts. Frege claims in several places after 1890 that sentences of significantly different grammatical structure can express the same thought and that they can do so in instances in which neither sentence is misleading as to the (or a) means of decomposing that thought. In 1892, Frege notes:

> It is thus not impossible that one way of analyzing a thought should make it appear as a singular judgment; another, as a particular judgment; and a third, as a universal judgment.[35]

Similarly, in 1906:

> [O]ne and the same thought can be split up in different ways and so can be seen as put together out of parts in different ways.[36]

The thesis of the multiple decomposability of thoughts seems in addition, though perhaps only sporadically in the mature theory, to play just the role played in *Grundlagen* by the multiple decomposability of contents, namely, that of underwriting an elucidatory equivalence between identity-claims regarding "new" objects and associated claims about objects and relations of a more-familiar kind. In an 1891 discussion of his all-important courses-of-value, Frege claims that the two sentences

$$(\mathbf{a})\ \forall x(x^2 - 4x = x(x-4))$$

and

$$(\mathbf{b})\ \dot{\varepsilon}(\varepsilon^2 - 4\varepsilon) = \dot{\alpha}(\alpha(\alpha - 4)),$$

the latter of which expresses the identity of the courses-of-values of the functions in question, express "the same sense, but in a different way."[37] Here we find the radical kind of re-carving seen in *Grundlagen*. And in an echo of the sentiment of *Grundlagen*, Frege says here:

> The possibility of regarding the equality holding generally between values of functions as a [particular] equality, viz. an equality between courses-of-value, is, I think, indemonstrable; it must be taken to be a fundamental law of logic.[38]

Michael Dummett takes it that Frege's multiple-decomposability view is untenable and particularly that it is not possible consistently to hold e.g. that (par) and (dir) express the same content while also holding that neither sentence is ontologically misleading.[39] As Dummett sees it, since (par) contains no terms for directions, it does not appear to "admit an analysis as a statement of identity between directions."[40] It can admit such an analysis, says Dummett, only if we take (par) to be a disguised way of expressing (dir), which is of course "precisely what Frege denies." The idea that (dir) is really an identity-sentence about directions, (par) a sentence concerning the relation between lines of being parallel, and that the two sentences express the same content, is, as Dummett sees it, inconsistent. Particularly problematic, in Dummett's view, is the idea that the recognition of a new decomposition of an already-grasped thought can lead to the grasp of a previously-unrecognized constituent of that thought.

Regarding pairs of sentences like (par) and (dir), and importantly including the pair

(eq) There are just as many *F*'s as *G*'s

and

(num) The number of *F*'s is the same as the number of *G*'s

Dummett remarks:

> It follows from Frege's criterion of identity of content that [these] pairs consist of two sentences expressing the same thought. And yet this conclusion runs counter to intuition. The reason is that someone could well understand the first sentence of [such a] pair without being in a position to understand the second. Of course, it in no way detracts from the claim that two sentences are synonymous that someone might understand one without understanding the other: he might simply not know the words. In [the cases under discussion], however, he might understand the first member of a pair without being as yet *capable* of grasping the thought expressed by the second, because he was not as yet in possession of a concept constitutive of the content of that second sentence: for instance, because he lacked the concept of a direction or of a cardinal number. That, indeed, is what Frege implies when he says that, by splitting up the content in the new way, we thereby attain a new concept. Yet this possibility is patently ruled out if the content of the one sentence is the same as that of the other: if it is the same, everyone who grasps the content of the one by that very act grasps that of the other, even if, through his ignorance of the words occurring in it, he does not know that it is the content of the other.[41]

But it is not clear that we should follow Dummett in this assessment.[42] The possibility Dummett takes to be "patently ruled out" is ruled out only if the grasp of a thought implies a grasp not just of the individual senses delivered by a particular, indicated decomposition of that thought, but also of all the senses delivered by all of its potential decompositions. And it is difficult to see why Frege should be thought to have held, or to have been required by reasons of coherence to hold, such a view. The view—that one can grasp a thought only if one recognizes the components of each of its potential decompositions—is indeed just what Frege denies when he claims that we can "acquire" new concepts by coming to see that a content is decomposable in a previously-unrecognized way.

The Fregean idea is that one can grasp a content or thought without yet having mastered the concepts or senses involved in each of its decompositions. Understanding "killed" and "Cato" (i.e., grasping their contents or senses), I can (and must) understand the whole content or sense expressed by "Cato killed Cato." Frege's idea is that I can do that before having grasped or taken any note

of the relation of *suicide*, though once I learn that each instance of "*x* committed suicide" expresses the thought or content expressed by the corresponding instance of "*x* killed *x*," I come to appreciate that new way of decomposing the already-grasped thought and, as Frege puts it, "obtain" a new concept, in this case that of suicide.[43] To "obtain" the concept of direction, similarly, on this picture, it suffices to come to appreciate a previously-unrecognized way of resolving a familiar content into constituents, a way indicated by the syntactic structure of the sentence (dir).[44]

Frege's commitment to the multiple-decomposability thesis survives the advent of the mature semantic theory; as noted, he clearly holds, after 1890, that thoughts are not uniquely structured. But the robustness of that thesis, and particularly the extent of the structural differences that separate different decompositions of a given whole, is less clear after the shift to the mature theory. As we've seen, Frege does occasionally say that two sentences, one of which expresses an identity between courses-of-value and the other of which expresses the coextensiveness of the associated functions, have the same sense; here we see the persistence with respect to thoughts of precisely the central, robust Grundlagen view about contents.[45] But Frege is, after 1890, generally more circumspect about the issue: in most discussions of the connection between course-of-value identities and coextensiveness assertions, he seems carefully to avoid the question of whether the sentences in question express the same thought.[46] Clearly maintained in the shift to the new theory is the idea that we "recognize" such objects as volumes, lengths, numbers, and most importantly extensions and courses-of-value, by recognizing an important equivalence between identity-statements regarding these objects and already-understood statements of related equivalences between pairs of, e.g., bodies, vectors, and concepts.[47] Also maintained is the idea that *some* pairs of structurally-different sentences express the same thought. Left somewhat unclear is Frege's commitment to the idea that the former sentence-pairs are instances of the latter, that is, that the important equivalence between the members of the pairs of sentences in question, that equivalence whose recognition gives rise to our acknowledgment of such objects as courses-of-value and numbers, is in fact the relation of expressing the same thought.

One important implication of the multiple-decomposability thesis, as Dummett points out, is that there is more to understanding a sentence than grasping the content or the thought it expresses. In coming to understand (dir) in the way Frege suggests, we need not only a grasp of its content—which we already have once we grasp the content of (par)—but also an appreciation of how to decompose that content in the new way, the way indicated by the structure of the new sentence (dir). As applied to thoughts, the thesis implies that two sentences might express the same thought despite the fact that it requires a non-trivial intellectual achievement to recognize that those sentences stand or fall

together. This picture, as Dummett points out, is in tension with a view about thoughts that Frege arguably also held, a view according to which thoughts are sufficiently finely individuated that no such intellectual achievement is required in order to recognize of two sentences that they express the same thought. We see once again a certain tension between the cognitive criterion for the individuation of thoughts and a more coarse-grained criterion that's arguably closely linked with some of the important work to which the theory of sense is put in Frege's writings.

2.4 Thoughts and Language Again

The fact that two sentences can differ significantly in overt grammatical structure without either of those structures proving misleading as to the composition of the senses expressed does not mean that overt grammatical structure is never in this way misleading.[48] As Frege points out, there is a sense in which the phrases "the number 20" and "every positive integer" play the same grammatical role, and hence in which "The number 20 can be represented as the sum of four squares" and "Every positive integer can be represented as the sum of four squares" have parallel superficial grammatical structure; but it would be incorrect to take the contents or the senses expressed by these sentences to be decomposable in structurally-similar ways.[49] Specifically, the noun phrases in each case play different roles, with one referring to an object and the other serving, as Frege puts it, to "confer generality," i.e. to quantify. Similarly, as Frege notes, the context of use can make even paradigmatic singular terms behave as quantifiers: in "The horse is a four-legged animal," what looks on the surface like a singular term is here standing as a quantifier, and neither its sense nor its reference is that of any singular term.[50] In general, as Frege sees it, the question of how the thought or content expressed by a sentence is composed out of parts, and hence the questions of how it can be so decomposed, and of how its value can be seen as determined by function-argument application, can be answered only on the basis of an understanding of, as one might intuitively put it, what that sentence says. To put it in terms of Frege's mature theory, these structural questions can often be answered only on the basis of a grasp of the thought expressed.

The sentences of Frege's formal language L_G of *Grundgesetze* are related to thoughts in much the same way as are the sentences of natural language, with some important exceptions. The central similarity is that each sentence of the formalism expresses a determinate thought. As with natural languages, the thought expressed by a sentence of L_G is determined by the senses of its parts and the order of their composition. Each of the sentential parts has both a sense and a reference, with the reference of complex pieces of language determined in

the usual function-argument way via the references of its parts; the reference of a complete sentence of L_G is its truth-value. That the sentences of L_G each express determinate thoughts marks Frege's formal language off significantly from our more-familiar modern formalisms, in which sentences are typically understood to be only partially interpreted, with fixed meanings assigned only to the logical particles. Frege's formal language is, by contrast, fully interpreted. To write down a series of formulas of L_G, as for example in presenting a proof, is to list a determinate series of thoughts.

The central differences between L_G and natural languages have to do with various requirements of expressive precision, requirements stemming from the intended role of L_G in the presentation of rigorous proofs. Frege explains his understanding of the role of proof as follows; here we repeat a passage quoted earlier:

> The aim of proof is, in fact, not merely to place the truth of a proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another. After we have convinced ourselves that a boulder is immovable, by trying unsuccessfully to move it, there remains the further question, what is it that supports it so securely? The further we pursue these enquiries, the fewer become the primitive truths to which we reduce everything; and this simplification is in itself a goal worth pursuing.[51]

This is of course just the kind of proof-theoretic reduction Frege has in mind for arithmetic: the goal is to demonstrate that the fundamental arithmetical truths, and so all truths of arithmetic, are provable from principles of pure logic.

As we saw in Chapter 1, Frege's decision to use a new formal language for his proofs, rather than ordinary mathematical German, is explained in *Begriffsschrift* in terms of the greater rigor afforded by the new language. Its use, says Frege, will help to "provide us with the most reliable test of the validity of a chain of inferences and to point out every presupposition that tries to sneak in unnoticed. . . ."[52] Similarly, as he puts it in *Grundlagen*:

> [E]very jump must be barred from our deductions. That [this requirement] is so hard to satisfy must be set down to the tediousness of proceeding step by step. Every proof which is even a little complicated threatens to become inordinately long. And moreover, the excessive variety of logical forms that have been developed in our language makes it difficult to isolate a set of modes of inference which is both sufficient to cope with all cases and easy to take in at a glance.
>
> To minimize these drawbacks, I invented my concept-script. It is designed to produce expressions which are shorter and easier to take

> in, and to be operated like a calculus by means of a small number of standard moves, so that no step is permitted which does not conform to the rules which are laid down once and for all. It is impossible, therefore, for any premise to creep into a proof without being noticed.[53]

And in *Grundgesetze*:

> Because there are no gaps in the chains of inference, every "axiom," every "assumption," "hypothesis," or whatever you wish to call it, upon which a proof is based is brought to light, and in this way we gain a basis upon which to judge the epistemological nature of the law that is proved. Of course the pronouncement is often made that arithmetic is merely a more highly developed logic; yet that remains disputable so long as transitions occur in the proofs that are not made according to acknowledged laws of logic, but seem rather to be based upon something known by intuition. Only if these transitions are split up into logically simple steps can we be persuaded that the root of the matter is logic alone.[54]

It is worth pointing out that the extreme care, and as even Frege himself puts it, the "tediousness" of presenting proofs in such a formal system is not a general requirement on the adequacy of proof. He does not claim that proofs must be presented in such a demanding formal system in order to be persuasive or to be up to the standards of justification required for ordinary mathematical demonstration:

> My purpose necessitates many departures from what is customary in mathematics. The requirements upon the rigor of proof inevitably entail greater length; anyone not bearing this in mind will indeed be surprised at how laboriously a proposition is often proved here that he believes he can grasp in one single act of understanding.
>
> . . .
>
> The length of a proof ought not to be measured by the yard. It is easy to make a proof look short on paper by skipping over many intermediate links in the chain of inference and merely indicating large parts of it. Generally people are satisfied if every step in the proof is evidently correct, and this is permissible if one merely wishes to be persuaded that the proposition to be proved is true. But if it is a matter of gaining an insight into the nature of this "being evident," this procedure does not suffice; we must put down all of the intermediate steps, that the full light of consciousness may fall upon them.[55]

If all we hope to establish is the *truth* of the conclusion in question, then all that's required is that each step in our proofs be clearly truth-preserving. We

don't, in this case, need to worry about the grounds on which we are implicitly relying when we present an inferential step as truth-preserving. Whether it's reliance on pure logic, on an evident truth grounded in intuition, or on a widely-acknowledged mathematical truth is of no concern in this case. But Frege's purposes are different: because he wants to establish exactly what's being appealed to in each proof, all such grounds must be explicitly acknowledged as premises. Every inferential step, in short, must clearly and obviously be made just via appeal to pure logic; everything else must be listed as a premise.

The two crucial features of the concept-script that make it better than natural language for the presentation of rigorous proofs are these.

First, the concept-script is expressively precise in a way that natural language is not: it has no ambiguous terms, no predicates with vague boundaries, and, importantly, no phrase whose grammatical category is misleading as to its logical role. In the concept-script, the fundamental logical categories as Frege understands them (e.g., singular term, first-level function-term of *n* arguments for each *n*, second-level function-term of *n* arguments for each *n*, quantifiers ranging over each of these) are each clearly marked in the syntax so that there is never any doubt about the logical structure of a given sentence. As Frege puts it:

> If it is one of the tasks of philosophy to break the domination of the word over the human spirit by laying bare the misconceptions that through the use of language often almost unavoidably arise concerning the relations between concepts and by freeing thought from that with which only the means of expression of ordinary language, constituted as they are, saddle it, then my concept-script, further developed for these purposes, can become a useful tool for the philosopher. . . . The mere invention of this concept-script has, it seems to me, advanced logic. I hope that logicians, if they do not allow themselves to be frightened off by an initial impression of strangeness, will not withhold their assent from the innovations that, by a necessity inherent in the subject matter itself, I was driven to make. These deviations from what is traditional find their justification in the fact that logic has hitherto always followed ordinary language and grammar too closely. In particular, I believe that the replacement of the concepts *subject* and *predicate* by *argument* and *function*, respectively, will stand the test of time. It is easy to see how regarding a content as a function of an argument leads to the formation of concepts. Furthermore, the demonstration of the connection between the meanings of the words *if, and, not, or, there is, some, all*, and so forth, deserves attention.[56]

In short, those grammatical distinctions found in natural language that have no bearing on logical inference are not reproduced in the concept-script, while all

aspects of content that are significant for logical inference are given explicit syntactic marking.

The second feature of Frege's concept-script that helps to guarantee that no premise can "creep into a proof without being noticed" is that its axioms and rules of inference are specified entirely syntactically.[57] As Frege puts it, the idea is that the concept-script can be "operated like a calculus by means of a small number of standard moves, so that no step is permitted which does not conform to the rules which are laid down once and for all."[58] That it is operable "like a calculus" is simply the point that no appeal to the contents of the formulas in question is needed for the verification that they conform to the rules of derivation that Frege has laid out in advance: the question of whether a given formula is a *Begriffsschrift* axiom or is immediately derivable via a *Begriffsschrift* rule from previously-encountered formulas is one that always has a clean, determinate answer, and one that can be decided just by examining the syntax of the formula.

The syntactic specification of the deductive system requires for its success the kinds of precision noted: the language must contain no ambiguous terms, no vague boundaries, and clear syntactic marking of logical categories. A further requirement imposed by the goal of a syntactic specification of truth-preserving rules is the requirement that every sentence have a determinate truth-value. This means that every singular term must have a determinate reference, and hence that every result of filling in some or all of the argument-places of a function-term must similarly have a determinate reference. Function-terms, including concept-terms, must in this sense be "totally defined," though see Chapter 3 for ways in which this requirement is weaker than might be supposed. No vague cases, and no "undefined" cases of the sort ubiquitous in natural language, can appear in L_G. In short, Frege's formal languages are "cleaned-up" versions of natural languages, with no referenceless, ambiguous, vague, or syntactically-misleading terms or sentences.

Because of the specific requirements on the formal languages he is to use in the course of proof, it is important to distinguish Frege's broad views about the workings of language in general from his narrower requirements on the language L_G. Some of the differences are clear: recall for example Frege's acknowledgment that pieces of ordinary language (e.g., "the least rapidly convergent series") can have sense without reference, and hence that a sentence can express a thought without having a truth-value. Frege holds that it's crucial to engineer L_G in such a way that this doesn't happen. Similarly, parts of natural language can be ambiguous; no parts of L_G can ever be. Natural languages can have pairs of terms that are interchangeable from the grammatical point of view ("every positive integer," "the number 20") but that play entirely different roles with respect to logical inference; L_G must never have such pairs. And so on. Other cases are less clear: as we'll see, Frege makes a number of pronouncements for example about the extent to which function- and-relation-terms can have incomplete boundaries,

pronouncements with respect to which it is less clear whether they apply to language in general, as Frege conceives of it, or merely to the artificially-constructed L_G. The reason this matters, for present purposes, is that the question of whether a certain imprecision is consistent with the expression of thought will turn out to be crucial to the question of whether there are in fact thoughts of a certain important mathematical kind. And the answer to this question will turn out to have implications for the question of how analysans- and analysandum-thoughts are taken by Frege to be related.

2.5 Where We Are

The truths of arithmetic, as understood in the terms of Frege's mature semantic theory, are a particular collection of thoughts. Similarly for the truths of logic, and indeed the truths of any science whatsoever. The question Frege seems to be most concerned to answer—namely, the question of whether the truths of arithmetic are grounded in pure logic—is the question of whether there is, for each truth of arithmetic, a *proof* from purely logical premises.

Frege uniformly claims that his proofs, those sketched in *Grundlagen* and given in detail in *Grundgesetze*, are proofs of arithmetical truths. He doesn't claim that they are proofs of the truths of some new and improved replacement science or of some theory interestingly related to arithmetic. He seems to take it that the truths in question are the thoughts expressed by such ordinary sentences as "every cardinal has a successor" and that they are the truths that have been investigated by mathematicians across centuries.[59]

But if the truths of arithmetic are expressed by the sentences of ordinary arithmetical discourse and also expressed by the final lines of Fregean derivations, then those final sentences must express the same thoughts as do the ordinary arithmetical sentences. And it is far from clear that they do so. Indeed, the large semantic gap between the sentences of ordinary arithmetic and the sentences found at the end of Frege's derivations might well be taken as reason to doubt that Frege was, in the end, interested in either the conceptual analysis of, or the foundations of, those truths entertained by ordinary people and mathematicians when they do arithmetic.

We turn in the next two chapters to the connection between the truths of ordinary arithmetic and the truths Frege takes himself to prove, with an eye toward the question of whether Frege can reasonably be taken at his word in his various discussions of conceptual analysis. We begin, in Chapter 3, with a potential difficulty for Frege's general semantic theory and particularly for the idea that both sentences of ordinary arithmetical discourse and those of the *Grundgesetze* formalism express determinate thoughts. In Chapter 4 we turn to the question of the relationship between the thoughts Frege proves and the thoughts of ordinary arithmetic.

Notes

1. [1892a] p 29/160.
2. [1879–91] p 6/6.
3. [1892d] p 115/105.
4. [1906a] p 401/332. See also Frege [1914] p 222/206: "What we prove is not a sentence, but a thought. And it is neither here nor there which language is used in giving the proof"; [1906a] p 387/318: "An inference simply does not belong to the realm of signs; rather, it is the pronouncement of a judgment made in accordance with logical laws on the basis of previously passed judgments. Each of the premises is a determinate thought recognized as true; and in the conclusion, too, a determinate thought is recognized as true."
5. *Begriffsschrift* (hereafter *Bff*) §2.
6. For the former view, see especially *Bff* §2. For the latter, note the discussion at *Bff* §8. Here we find also the third (and most explicit) case of the contents of individual signs: in identity-contexts, says Frege, the contents of the signs flanking the identity-sign are the signs themselves rather than their ordinary contents (which are here characterized not as ideas but as objects). As has often been noted, the argument Frege gives here for the "bifurcation" of content is later [1892a] turned into an argument for the distinction between sense and reference.
7. *Bff* §2; translation slightly amended.
8. [1882a] p 97–98/90–91: "I did not wish to present an abstract logic in formulas, but to express a content through written symbols in a more precise and perspicuous way than is possible with words."
9. *Bff* §3.
10. For discussion of this issue, see Kremer [2010].
11. The three papers are [1891], [1892a], and [1892c].
12. In an amusing indication of his care with respect even to throwaway examples, Frege corrects himself here to note that perhaps what's meant is really "all properly constituted horses are four-legged animals" [1892c] p 196/185.
13. [1892c] p 196/185 note.
14. Similarly, in the previous year, Frege writes that two sentences express different thoughts when they intuitively "tell us quite different things" despite the fact that their parts refer to the same objects and functions. See [1891] p 14/145.
15. [1891] p 14/145. See also [1892a] p 32/162, in which Frege argues that the sentences "the morning star is a body illuminated by the sun" and "the evening star is a body illuminated by the sun" must express different thoughts since "[a]nybody who did not know that the evening star is the morning star might hold the one thought to be true, the other false."
16. [1906b] pp 213–214/197–198.
17. What counts as "significantly different" turns on the judgments in question; what's important is that the eras involve significant conceptual development with respect to the subject-matter of those judgments.
18. See [1892a] p 29/160, [1892c] p 196/185 note. Also [1897a] p 142/131: "The sense of an assertoric sentence I call a thought. Examples of thoughts are laws of nature, mathematical laws, historical facts. . . ."
19. *Bff* Preface p iv/6.
20. Letter to Husserl, December 9, 1906; [1980] pp 70–71; translation slightly amended.
21. There is at least an appearance of circularity here, since for Frege it is thoughts and not sentences that bear logical relations to one another, with the result that it might in principle be necessary to determine whether A and B express the same thought prior to determining whether they express logical equivalents. But not necessarily, and this for two reasons. First, with respect to the sentences of a formal language, one may have good reason (e.g., via derivation in a formal system known to be reliable) to hold that A and B express logical equivalents prior to knowing anything further about whether they express the same thought. Second, more generally, the question of whether what's expressed by a

given sentence logically implies what's expressed by another, and vice-versa, can in some cases clearly be determined just via an ordinary understanding of those sentences, without answering the question of whether the sentences are similar enough to count as expressing the same thought.

22. See e.g. [1891] p 13/145, in which Frege notes that "$2^4 = 4^2$" and "$4 \cdot 4 = 4^2$" express different thoughts; also *Grundgesetze* (hereafter *Gg*) 1 §2, in which "$2^2 = 4$" and "$2 + 2 = 4$" are said to have different senses.
23. See, e.g., [1891] p 6/140: "[A] function by itself must be called incomplete, in need of supplementation, or unsaturated (*ungesättigt*)." Also *Gg* 1 §1.
24. ". . . not all the parts of a thought can be complete; at least one must be 'unsaturated' or predicative; otherwise they would not hold together" ([1892c], p 205/193).
25. As to the question of whether this function is defined on all objects, see Chapter 3.
26. [1892a] p 27/158.
27. See e.g. [1892a] pp 26–27/158.
28. See Gareth Evans [1981] p 283: "[T]he sense of a singular term is a way of thinking about a particular object; something that obviously could not exist if that object did not exist to be thought about." Evans's [1982] extended treatment in *The Varieties of Reference* (Chapter 1) is more nuanced. Here Evans notes the texts in which Frege appears committed to the existence of terms lacking reference while possessing sense, but takes it that this commitment is both tenuous and mistaken and that its denial is more in keeping with Frege's central views. As Evans ([1982] p 22) puts it:

> Surely the choice of entities of a given type as appropriate semantic values for expressions of a given category must mean that the functioning of any member of that category as a viable element of language depends upon its being associated with such an entity.

Hence, as Evans sees it, one can't coherently take truth-values to be (the) appropriate semantic values for sentences while holding that some intelligible sentences lack them. And given the compositionality of reference and of sense, one therefore can't take individual terms to be intelligible (i.e., to express a sense) while lacking a reference.

But because Frege never explains the intelligibility of a sentence or of its components in terms of their possession of reference, it is unclear why one should follow Evans in taking Frege to be committed to such a link between intelligibility and reference. The intelligibility of linguistic items in general is indeed tied on Frege's view to their role in the expression of truth-evaluable contents or thoughts. But this does not rule out the intelligible occurrence of those items in combinations which, given their general role, are guaranteed to determine no truth-value at all in virtue of the failure of the normal conditions necessary to their use in the expression of a truth or a falsehood. Just such failure is evident in the Fregean examples of non-referring definite descriptions ("the least rapidly convergent series") or empty names ("Odysseus").

29. [1892a] p 28/159. See also [1892a] p 41/169, where Frege notes the "imperfection of language, from which even the symbolic language of mathematical analysis is not altogether free; even there combinations of symbols can occur that seem to refer to something but (at least so far) do not mean anything, e.g. divergent infinite series."
30. From [1892a] p 32/162:

"Is it possible that a sentence as a whole has only a sense, but no reference? At any rate, one might expect that such sentences occur, just as there are parts of sentences having sense but no reference. And sentences which contain proper names without reference will be of this kind. The sentence 'Odysseus was set ashore at Ithaca while sound asleep' obviously has a sense. But since it is doubtful whether the name 'Odysseus,' occurring therein, refers to anything, it is also doubtful whether the whole sentence does."

See also [1892a] p 33/163 note: "It would be desirable to have a special term for signs having only sense. If we name them, say, representations, the words of the actors on the stage would be representations. . . ."

31. [1892a] p 41/169. Also [1906b] p 214/198: "In myth and fiction thoughts occur that are neither true nor false. Logic has nothing to do with these. In logic it holds good that every thought is either true or false, *tertium non datur*."
32. *Grundlagen* (hereafter *Gl*) §64.
33. In the early semantic theory, Frege often writes as if the "contents" expressed by declarative sentences are composed out of, or determined by, the "contents" of their parts, and further that the content of a singular term is the object later known as its reference, and the content of a predicative phrase something like what is later known as the function or concept it refers to. But that Frege has a well-worked-out and univocal view on these issues is not well supported by the texts.
34. *Gl* §64.
35. [1892c] p 200/188–189.
36. [1906b] p 218/201–202.
37. [1891] p 10–11/143. Note the absence of a modern universal quantifier in the first of these two sentences as Frege writes it; the quantification is for him given by the style of lettering.
38. [1891] p 10/142.
39. Dummett [1991a] p 169.
40. Ibid.
41. Dummett [1991a] p 172.
42. For discussion of difficulties with Dummett's account here, see Levine [2002].
43. See the passage from *Bff* §9 quoted above. Dummett [1991a] p 174 discusses this example, concluding that the sense of "Cato committed suicide" is more complex than that of "Cato killed Cato," since its apprehension requires the understanding of the linguistic device "suicide." Here again we find the line of argument questioned above; Dummett takes it that Frege is committed to the idea that "a grasp of the sense of a sentence containing a word (e.g. "suicide") . . . requires an understanding of that word." But it is difficult to see why Frege should be thought to hold this view: one of course must grasp the sense of that word in order to understand a *sentence* containing the word, but that the sense of that sentence can be expressed by a structurally very-different sentence not containing the word is, from the Fregean point of view, reason enough to conclude that one can grasp that thought without grasping the sense of that word.
44. It does, to be sure, sound odd to say that one can grasp the content of (dir) without having the concept of *direction*. But if Frege is right, there is nothing odd about the claim at all, but only about this way of putting it: normally, we characterize the content that a person is said to grasp by using a sentence that he or she might use or a structurally-similar translation thereof, and this convention is flouted when we characterize a person's grasp of the content in question by adverting to (dir), when that person would use (par) or a structural equivalent.
45. [1891] p 11/143.
46. See, e.g., *Gg* 1 §§3, 9, 20.
47. See, e.g., *Gg* 1 §3, §9; *Gg* 2 §58 note; [1891] p 10/142.
48. That the structure of ordinary language can be significantly misleading with respect to logical structure, and that a primary virtue of his formal languages is the more faithful reflection of that structure in the syntax of sentences, is a theme Frege repeats early and late. See *Bff* Preface p vi/7; *Bff* §3; [1882b] pp 84–85; [1897a] p 154/142.
49. See *Bff* §9.
50. [1892c] p 196/185 note.
51. *Gl* §2.
52. *Bff* Preface p iv/6.
53. *Gl* §91.
54. *Gg* 1 Inroduction p vii/3
55. *Gg* 1 Introduction pp vii–viii/4–5.
56. *Bff* Preface p vii/7. The remark about subject and predicate is explained several paragraphs later as follows in *Bff* §3:

> A distinction between *subject* and *predicate* does *not occur* in my way of representing a judgment. In order to justify this I remark that the contents of two judgments may differ in two ways: either the consequences derivable from the first, when it is combined with certain other judgments, always follow from the second, when it is combined with these same judgments, or this is not the case. The two sentences "The Greeks defeated the Persians at Plataea" and "The Persians were defeated by the Greeks at Plataea" differ in the first way. Even if one can detect a slight difference in meaning, the agreement outweighs it. Now I call that part of the content that is the *same* in both the *conceptual content*. Since *it alone* is of significance for our concept-script, we need not introduce any distinction between sentences having the same conceptual content. . . . Now, all those peculiarities of ordinary language that result only from the interaction of speaker and listener—as when, for example, the speaker takes the expectations of the listener into account and seeks to put them on the right track even before the complete sentence is enunciated—have nothing that answers to them in my formula language, since in a judgment I consider only that which influences its *possible consequences*. Everything necessary for a correct inference is expressed in full, but what is not necessary is generally not indicated. . . .

57. That is, Frege gives an entirely syntactic specification of axiom-*sentences*, each of which is guaranteed to express only a logical truth, and indeed one whose status as a logical truth is (Frege thought) easy to see. Similarly for rules of inference.
58. *Gl* §91.
59. Note, e.g., the opening of *Gg* 1 §0:

> In my *Grundlagen der Arithmetik*, I sought to make it plausible that arithmetic is a branch of logic and need not borrow any ground of proof whatever from either experience or intuition. In the present book this shall now be confirmed, by the derivation of the simplest laws of Numbers by logical means alone. But for this to be convincing, considerably higher demands must be placed on the conduct of proof than is customary in arithmetic.

Also note Frege's own natural-language renderings of the important truths proven in *Grundgesetze*, e.g. the rendering (in *Gg* 1, Appendix 3, "Table of Important Theorems") of Theorem 110 as "The number one follows immediately after the number zero." Similarly, note the natural-language renderings discussed above of the theorems proven in [1881] and of the truths whose proofs are sketched in *Gl* §§78–79. That the arithmetical truths in question are just those truths to which we ordinarily appeal in the physical sciences is brought out at *Grundlagen* §87, where Frege sums up what he hopes to have accomplished by 1884 as follows:

> I hope I may claim in the present work to have made it probable that the laws of arithmetic are analytic judgments and consequently a priori. Arithmetic thus becomes simply a development of logic, and every proposition of arithmetic a law of logic, albeit a derivative one. To apply arithmetic in the physical sciences is to bring logic to bear on observed facts. . . .

3

Thoughts and Sharp Boundaries

3.1 The Issue

Frege's view as presented so far includes the theses (1) that ordinary arithmetical discourse involves the expression of true and false thoughts and (2) that some among these thoughts are to be demonstrated by the logicist reduction to be grounded in pure logic. Also included is the claim (3) that the final sentences of Frege's *Grundgesetze* derivations are themselves intended to express thoughts, ones whose logical grounding is immediately established by those derivations. The question raised in Chapter 2, and to which we'll return in Chapter 4, is that of how Frege conceives the relationship between the thoughts just mentioned in (3) and those mentioned in (1).

Before returning to that question, we digress here to treat a potential difficulty that arises for thesis (1) and thus for the outline just sketched of Frege's logicist strategy.[1] The difficulty is this: Frege claims in many places that functions, and hence concepts, must have "sharp boundaries," by which he seems to mean that they are *total*, in the sense of delivering a value for every argument (or *n*-tuple of arguments) of appropriate logical type. There seems good evidence, additionally, that he holds the following views: that a function-term that fails to refer to a function thereby fails to refer at all, and that, as a consequence, no sentence in which that function-term appears can have any reference, which is to say that it can have no truth-value.[2] If the sentences of ordinary arithmetic are to express truths and falsehoods, then, it is essential, on this view, that their function-terms refer to total functions. The difficulty is that they don't seem to do so. The sense and reference of such function-terms as "+," in ordinary use, are presumably determined largely by our use of them together with relevant mathematical facts. And no such considerations determine what value the reference of such a term should have when applied to the pair of arguments <7, the Eiffel Tower>.

It is worth emphasizing here that the difficulty is not just that outlying sentences like "7 + the Eiffel Tower < 14" have no truth-value; the problem is that, given the constellation of theses just outlined, it follows that "7 + 3 = 10," similarly,

has no truth-value, since its function-term fails to refer to a total function and hence fails to refer. All such ordinary sentences fail, then, on this view, to express truth-evaluable thoughts.

We'll begin by looking at the "sharp boundaries" passages in question and at the implications for Frege's project of reading them in the way just suggested, i.e. as demanding that all functions be total. In the end, I'll argue that, despite the textual evidence for this reading, Frege cannot be read as holding that all functions are total. What he does mean in the "sharp-boundaries" passages is significant for our understanding of how he understands the scope of theories and of logic. Most importantly for present purposes, it's not the case that these passages undermine the view of ordinary arithmetical discourse as consisting of the expression of true and false thoughts.

3.2 The Texts

The sharp-boundaries texts include these:

An 1896 claim that occurs in the course of a criticism of Peano:

> [E]very concept must have sharp boundaries, so that it is determined for every object whether it falls under the concept or not.[3]

A discussion in *Grundgesetze* that applies the requirement to functions more generally:

> [A] first-level function of one argument must always be such as to yield an object as its value, whatever object we may take as its argument. . . .[4]

> [E]very first-level function of two arguments must have an object as its value for any one object as its first argument and any other object as its second. . . .[5]

Similarly for functions of higher level.[6]
An 1891 insistence on the

> . . . requirement, as regards concepts, that, for any argument, they shall have a truth-value as their value; that it shall be determinate, for any object, whether it falls under that concept or not.[7]

And there are more.[8]

If Frege means that, strictly speaking, all functions and concepts must be defined everywhere, even on arguments foreign to the subject matter under

discussion, then this requirement undermines a good deal of the picture one might have had of Frege's understanding of language, given his frequent examples of (apparently) truth-evaluable thoughts expressed by ordinary sentences.[9] It's not just e.g. "+" as ordinarily understood that fails to refer to a total function. Predicates and relation-terms whose natural homes are in the physical sciences, for example, are not generally defined over arbitrary objects outside of this range. And even without looking at "outlying" arguments, troubles are easily found close to home: the phrase "the eldest child of . . .," as applied to a childless adult, gives no determinate answer and so must too, on the requirement of totality, count as reference-less. No sentence employing this function-term can express a true or false thought on the view in question. In short, if Frege holds that a condition of reference for function-terms is that they refer to total functions, then he cannot coherently hold that ordinary discourse or ordinary scientific inquiry involves, in his words, a "common stock of thoughts," at least not of true or false ones.[10]

On one important interpretation of Frege's work, this is just the conclusion that Frege endorses. As Joan Weiner understands the project, Frege was not, after all, concerned with the functioning of language in ordinary, including ordinary scientific, discourse. His theory of sense and reference, on this view, is meant to apply only to languages expressly designed for rigorous scientific inquiry and proof, and meeting some very high standards of completeness. Thought-expression is a feature, in short, just of languages as careful and systematic as is that of *Grundgesetze*. As Weiner puts it:

> [G]iven [Frege's] requirement that each predicate pick out a concept with a sharp boundary, few, if any, of our everyday sentences have comprehensible sense or truth-values.[11]

This means that whatever mathematicians were doing prior to the systematization of arithmetic given by Frege's work, they were not expressing, debating, or proving arithmetical truths. The sentences they used expressed neither truths nor falsehoods. And the same must be said for us: since the final systematization of arithmetic has not yet been accomplished, and since our arithmetical education does not inculcate a grasp of senses that determine total functions, we cannot be said to be engaged in expressing, debating, or proving arithmetical truths. Indeed, as Weiner acknowledges, it is a consequence of her interpretation of Frege that, on the Fregean view, "it is not even obvious that we know any truths of arithmetic."[12] For what we know, on the Fregean picture, is what's expressed by some of the sentences we affirm, and none of these, on that picture, express truths.

One difficulty for this view is that most of Frege's discussions of thoughts and their expression in language have to do not with the formulas of a precisely

specified formal system, but with ordinary sentences of natural language. In his most pointed discussions of sense and reference, e.g. in the trio of papers from 1891–92 and in the later "Der Gedanke" and related essays, he never claims that the sentences of ordinary discourse fail to express thoughts.[13] And in his discussions of the ways in which rigorous formal languages offer improvements on ordinary language, he never claims that a fault of ordinary sentences is that they fail to express thoughts. Ordinary language is sloppy, to be sure, on Frege's view, in ways already touched upon above, but that the sloppiness is sufficient to undermine the attempt to use the sentences of such a language to express truths and falsehoods is never cited by Frege as one of its difficulties.

A more serious problem for the view that only formal languages like that of *Grundgesetze* are capable of expressing true and false thoughts, in virtue of the total definition of the functions referred to in such a language, is that the functions referred to in *Grundgesetze* are very clearly not total. Consider Frege's treatment of course-of-values names. At *Grundgesetze* 1 §10, in the course of laying down the references of his terms, Frege discusses the fact that course-of-values names have not yet been assigned determinate reference via the stipulations given to this point. Particularly instructive is his means of remedying this defect.[14] The discussion is as follows:

> Although we have laid it down that the combination of signs
>
> $$\text{“}\dot{\varepsilon}\,\Phi(\varepsilon) = \dot{\alpha}\,\Psi(\alpha)\text{”}$$
>
> has the same denotation as
>
> $$\text{“}\forall x\,(\Phi(x) = \Psi(x)),\text{”}$$
>
> this by no means fixes completely the denotation of a name like "$\dot{\epsilon}\Phi(\epsilon)$." We have only a means of always recognizing a course-of-values if it is designated by a name like "$\dot{\epsilon}\Phi(\epsilon)$," by which it is already recognizable as a course-of-values. But we can neither decide, so far, whether an object is a course-of-values that is not given us as such, and to what function it may correspond, nor decide in general whether a given course-of-values has a given property unless we know that this property is connected with a property of the corresponding function. . . .
>
> How may this indefiniteness be overcome? By its being determined for every function when it is introduced, what values it takes on for courses-of-values as arguments, just as for all other arguments. Let us do this for the following functions considered up to this point.[15]

Here Frege lists the three functions considered up to now, and notes that two of them are reducible to the third, i.e. to the identity-function:

> Since in this way everything reduces to consideration of the function $\xi = \zeta$, we ask what value this has if a course-of-values occurs as argument. Since up to now we have introduced only the truth-values and courses-of-values as objects, it can only be a question of whether one of the truth-values can perhaps be a course-of-values.[16]

Consider a particular course-of-values, say, $\grave{\epsilon}(\epsilon = \epsilon)$. The indeterminacy in question is, with respect to this case, the fact that sentences of the form "___ = $\grave{\epsilon}(\epsilon = \epsilon)$" and "$\grave{\epsilon}(\epsilon = \epsilon)$ = ___" have no truth-value in cases in which the blank is filled in by anything other than a term of the form "$\grave{\alpha}\Phi(\alpha)$." One might think that the problematic indeterminacy of reference on the part of "$\grave{\epsilon}(\epsilon = \epsilon)$" is due to the fact that a vast number of identity-questions regarding its purported reference have been left unanswered. But that this would be a misleading way of construing the difficulty is made clear by Frege's means of addressing it. Frege does not solve the problem by providing a means of answering those unanswered questions in general. He does not bring it about that the concept-phrase "___ = $\grave{\epsilon}(\epsilon = \epsilon)$" determinately holds or fails to hold of each object. He instead brings it about that every way of completing this concept-phrase with a singular term of the language, i.e. every sentence formed in this way, has a determinate truth-value. As he says, since the only objects that have been "introduced" up to this point (i.e., the only ones to which one can refer in the language) are truth-values and courses-of-value, and since the identity-questions about courses-of-value have been settled already, the only question left to be settled is that of whether the concept-phrase "___ = $\grave{\epsilon}(\epsilon = \epsilon)$" is satisfied by terms for either of the truth-values. Similarly for each of the course-of-value terms: for each such term $\grave{\epsilon}(\Phi(\epsilon))$, it must be determined whether "___ = $\grave{\epsilon}(\Phi(\epsilon))$" is satisfied by terms for either of the truth values.

Frege settles the issue by a simple arbitrary stipulation: "[L]et us lay it down that $\grave{\epsilon}(—\epsilon)$ is to be the True and $\grave{\epsilon}(\epsilon = \sim(\forall x)x = x)$ is to be the False."[17] This neatly solves the problem: each of the problematic sentences now has a determinate truth-value. The stipulation does not, of course, have the result that the concept-phrases of the form "___ = $\grave{\epsilon}(\Phi(\epsilon))$" refer to total concepts: nothing has been said to determine whether any such phrase is satisfied by objects not named in the language. Similarly, nothing has been said to determine whether the courses-of-values in question fall under or fail to fall under concepts not yet referred to in the language. Frege's remarks on this circumstance are simply that:

> With this we have determined the courses-of-values so far as is here possible. As soon as there is a further question of introducing a function that is not completely reducible to functions known already, we can stipulate what value it is to have for courses-of-values as arguments; and this can then be regarded as much as a further determination of the courses-of-values as of that function.[18]

The fact that the concepts referred to by *Grundgesetze* concept-expressions are very clearly not total does not, as Frege sees it, stand in the way of the entirely unproblematic functioning of the language of *Grundgesetze*. Indeed, the failure of totality does not even stand in the way of the expression of determinate, truth-evaluable thoughts by the formulas of that language. As Frege puts it in *Grundgesetze* 1 §32, having argued that each of the ways of combining function- and argument-expressions of the language has a well-defined reference (in the sense that no formula constructed out of them lacks a truth-value):

> In this way it is shown that our eight primitive names have denotation, and thereby that the same holds good for all names correctly compounded out of these. However, not only a denotation, but also a sense, appertains to all names correctly formed from our signs. Every such name of a truth-value *expresses* a sense, a *thought*. Namely, by our stipulations it is determined under what conditions the name denotes the True. The sense of this name—the *thought*—is the thought that these conditions are fulfilled.
>
> . . .
>
> [W]e have in every correctly-formed proposition of the concept-script [here Frege refers to a sentence prefixed with the judgment-stroke] a judgment that a thought is true; and here a thought certainly cannot be lacking.[19]

In short, the sentences of *Grundgesetze* express unproblematic thoughts, ones that have determinate truth-values. This despite the fact that the functions referred to via their predicative parts are all very partial, defined only over an extremely limited collection of objects.

What, then, does Frege mean by the sharp-boundaries requirement, given that neither ordinary (arithmetical) discourse nor the extremely careful language of *Grundgesetze* come close to meeting the requirement of totality?

Here it will be important to distinguish the requirement of totality from what we'll call the requirement of *linguistic completeness*. The latter is the requirement for a given language that each of its well-formed concatenations of symbols has a determinate reference. This is a requirement that Frege takes to be crucial for languages that are intended for the expression of rigorous arguments, and especially for those like his own concept-script, with respect to which the conditions of rigorous proof are to be laid down syntactically. For the presence in a language of well-formed phrases lacking reference is an invitation to fallacious reasoning. As Frege puts it:

> It seems to be demanded by scientific rigor that we should have provisos against an expression's possibly coming to have no reference; we

> must see to it that we never perform calculations with empty signs in the belief that we are dealing with objects. People have in the past carried out invalid procedures with divergent infinite series.[20]

Not just mathematical calculations, but ordinary reasoning as well can be derailed by the presence of non-referring parts of language. Even the law of excluded middle, as Frege points out, will lead us astray if we rely on it in contexts in which the application of a concept-expression to an object-name gives no definite result. Particularly important to Frege, given his interest in induction, are the Sorites-style fallacies that are invited by the use of concept-expressions that fail to deliver a truth-value on some completions by object-names.[21] In short, a language that fails the requirement of linguistic completeness is one that's not suited for maximum argumentative clarity and rigor. Frege is explicit in his application of the requirement of linguistic completeness to the language of *Grundgesetze*, going so far as to present a detailed proof that the requirement is satisfied.[22]

As to the relationship between the requirement of linguistic completeness and the requirement of totality: if a given theory or discourse is concerned just with a specific range of objects—say, the integers—then the language used there will satisfy the requirement of linguistic completeness as long as its function-expressions refer to functions defined on all of the arguments from this range. The requirement of linguistic completeness for such a language is considerably weaker than the requirement that its function-terms refer to total functions, i.e. to ones whose domain includes outlying objects.

Frege clearly has just such restricted languages in mind much of the time. The central example of this is just the language of *Grundgesetze* itself. With respect to the context of ordinary arithmetic, additionally, Frege notes in "Function and Concept" that:

> So long as the only objects dealt with in arithmetic are the integers, the letters *a* and *b* in "*a* + *b*" indicate only integers; the plus-sign need be defined only between integers.[23]

These two examples are of course counterexamples to the requirement of totality, and further evidence that Frege's sharp-boundaries requirement is not in fact the strong totality-requirement.

What, then, does Frege mean when he claims that functions must be defined for "all arguments"? Here, the evidence points to the view that by "all arguments," Frege means to include all of those arguments in the domain of the science or theory at hand. Frege's real interest, I'll suggest, is the requirement of linguistic completeness, a requirement in which he has a clear vested interest. Consider the first sharp-boundaries passage quoted above, together with a little more of its context:

> We can also argue for this requirement on the ground that the law of the excluded middle must hold. For that implies that every concept must have sharp boundaries, so that it is determined for every object whether it falls under the concept or not.[24]

The "requirement" to which Frege refers is spelled out in the previous paragraph, which consists largely in a criticism of Peano for having failed to define his function-terms "num" and "ϵ" in such a way that every result of completing their argument-places with names from Peano's language will have a determinate reference:

> [I]f, in our example, we take as before "then *u* is not a class" as a consequent, we must also understand by *u* something that is not a class, and hence in this case too "num *u*" must *refer to* something, if it is to be possible to judge whether the condition "if num *u* does not coincide with num *v*" is satisfied, and equally it must also be possible to judge whether a given relation maps *u* into *v* or *v* into *u* for the case where *u* is not a class. So the definitions of "num *u*" and of mapping ("$f \epsilon vfu$") must be formulated accordingly. Since one cannot know from the outset in which sentences these signs will occur and what restrictions will be thereby placed on the meanings of the letters, the definitions are to be constructed in such a way that a reference is guaranteed these combinations of signs for every reference of the letters.[25]

That is: the requirement in question is the all-important requirement of linguistic completeness, not the requirement that the function-signs be defined over outlying objects.

Similarly for the *Grundgesetze*'s sharp-boundaries passages quoted above. Despite his talk of functions defined over "any" arguments, Frege's practice makes it clear that the only objects meant here are those that have names in the language of *Grundgesetze*, i.e. courses-of-value and truth-values. Once again, the only coherent way to read these passages is as insisting not on the strong requirement of totality for functions, but on the requirement of linguistic completeness.

3.3 Piecemeal Definition and New Objects

Perhaps the most important context in which Frege discusses what looks like the requirement of totality, certainly the context in which the sharp-boundaries requirement is raised in the most sustained and most forceful way, concerns Frege's prohibition against what he calls "piecemeal definition." What he means

by this, and the reasons for prohibiting it in a good scientific system, are laid out in *Grundgesetze* as follows:

> A definition of a concept (of a possible predicate) must be complete; it must unambiguously determine, as regards any object, whether or not it falls under the concept (whether or not the predicate is truly ascribable to it). . . . We may express this metaphorically as follows: the concept must have a sharp boundary. . . . [A] concept that is not sharply defined is wrongly termed a concept. Such quasi-conceptual constructions cannot be recognized as concepts by logic; it is impossible to lay down precise laws for them. The law of excluded middle is really just another form of the requirement that the concept should have a sharp boundary.
>
> Now from this it follows that the mathematicians' favorite procedure, piecemeal definition, is inadmissible. The procedure is this: First they give the definition for a particular case—e.g. for positive integers—and make use of it; then, many theorems later, there follows a second definition for another case—e.g. for negative integers and zero. Here they often commit the further mistake of making specifications all over again for the case they have already dealt with. Even if in fact they avoid contradictions, in principle their method does not rule them out. What is more, as a rule they do not attain to completeness, but leave over some cases, as to which they make no specification; and many are naïve enough to employ the word or symbol for these cases too, as if they had assigned a meaning to it. . . . But the chief mistake is that they are already using the symbol or word in theorems before it has been completely defined—often, indeed, with a view to further development of the definition itself. So long as it is not completely defined, or known in some other way, what a word or symbol means, it may not be used in an exact science—least of all with a view to further development of its own definition.[26]

The dangers associated with piecemeal definition, then, include (1) the possibility of introducing a contradiction by providing later stipulations incompatible with earlier ones, and (2) the possibility of failing to cover some cases, with the result that some combinations of symbols are left without reference.

But if piecemeal definition is ruled out, then what are we to say of those episodes in the history of mathematics in which familiar function-signs have had to have their reference extended so as to be applicable to newly-introduced objects? Frege's answer is as follows:

> Now, of course it must be admitted that scientific progress, which has been effected by conquering wider and wider domains of numbers,

> made such a procedure almost inevitably necessary; and this necessity might serve as an excuse. [Here appears a lengthy footnote directed at Peano.] It would indeed have been possible to replace the old symbols and terms by new ones, and logic really demands this; but that is a hard decision to make. And this horror over the introduction of new symbols or words is the cause of many obscurities in mathematics. The old definitions likewise could have been rejected as unsound, and new ones used, in order to set up the science over again from the beginning; but such a clean cut was never made . . . In this way people have got used to piecemeal definition; and what was originally an awkward makeshift became customary, and was admitted as one of the legitimate methods of science. The result is that nowadays hardly anybody is shocked when a symbol is first defined for a limited domain and then used in order to define the same symbol once more for a wider domain; for general custom has a power of justifying what is done, just as fashion can give the cachet of beauty to the most detestable mode. It is all the more necessary to emphasize that logic cannot recognize as concepts quasi-conceptual constructions that are still fluid and have not yet been given definitive and sharp boundaries, and that therefore logic must reject all piecemeal definition.[27]

The proper procedure to use when extending the mathematical domain to new objects, as Frege sees it, is simply to use a new sign for the newly expanded function. This is, as he puts it, "what logic really demands." We leave e.g. the old addition sign behind, moving to a new sign defined all at once for the whole of the new domain of objects. We in this way entirely avoid the central dangers associated with piecemeal definition, those of introducing inconsistency via overlapping stipulations and of leaving some combinations of signs meaningless via gappy stipulations.

Notice that the difficulty with which Frege is concerned when new mathematical objects are introduced is a difficulty regarding the definitions of the function-terms of the newly expanded science: the difficulty is that, if we simply add new stipulations to the previously recognized definitions of the original function-symbols, we run the danger of contradicting ourselves or of leaving some combinations of symbols out of account. We run the danger, in short, of bringing into our system new combinations of symbols that have no determinate reference. The difficulty pointed out by the introduction of new elements to mathematics is importantly *not* what one might have expected, given some of the ways in which Frege presents the condition of sharp boundaries. The difficulty is not that, having introduced the new elements, we are forced to recognize a logical flaw in the old science—the flaw, namely, that its function-terms refer only to partial functions, ones not defined on the newly recognized objects.

Frege makes no such claim. The problematic system is the one obtained by flawed (inconsistent or gappy) expansion of the old system, and not the old, partial system itself.

The repeated expansion of the domain of mathematical objects is as Frege sees it a series of episodes in which mathematicians recognize the existence of objects about which they had previously been ignorant—but which had, nevertheless, always existed. Mathematical activity is a process of discovery, not of creation. At each historical stage, then, the functions referred to in mathematics are only partial, having no determinate values for the as-yet undiscovered objects. Frege's view about this circumstance is not that there is anything wrong with these partial functions or that logic must eschew them. His claim is rather the considerably more moderate one that the introduction of new objects into the realm of arguments under consideration must, in the interests of logical rigor, be accompanied not by piecemeal add-ons to the definitions of function-signs already in use but rather by the introduction of new function-signs defined in one fell swoop over all of the objects of the new domain:

> [I]t is so easy to avoid a plurality of definitions for one and the same symbol. Instead of first defining a symbol for a limited domain and then using it for the purpose of defining itself in regard to a wider domain, we need only choose different signs, and confine the meaning of the first, once for all, to the narrower domain. . . .[28]

Similarly, as Frege puts it in a 1914 lecture course, if Carnap's notes are to be trusted:

> In the development of mathematics one does, however reach certain points where one wants to expand the system. But then one has to begin from scratch again. In any case, there always has to be a complete *system* at hand that is logically unproblematic. E.g. one would have to proceed as follows: as long as the plus sign + is used only for positive whole numbers, one chooses a different sign for it, e.g., ¬.[29]

The Fregean requirement is, as we might put it, that the functions involved in a given science must be defined on all of the arguments in the domain of that science. And the reason for this is the straightforward requirement of linguistic completeness.

Some of Frege's more puzzling pronouncements regarding the sharp-boundaries requirement seem to conflict with this picture and to involve the claim that the function signs of mathematics must be defined over clearly nonmathematical objects. Thus, for example, in *Grundgesetze* 2 §§63–64, Frege seems to claim, contrary to his position in "Function and Concept," that the addition sign must be defined over the

moon, and can't be restricted to a particular domain, e.g. that of the numbers. But it's important to see that this is not his point. The passage reads as follows:

> [A] first-level function of one argument must always be such as to yield an object as value, whatever object we may take as its argument—whatever object we may use to "saturate" the function.
>
> We must make the corresponding requirement as regards functions with two arguments. The expression
>
> "the sum of one object and another object"
>
> purports to be the name of such a function. Here too, then, it must be determinate, as regards any one object and any other object, which object is the sum of the one and the other; and there must always be such an object. If that is not the case, then it is likewise indeterminate which object gives the result one when added to itself. In that case, therefore, the words "something that gives the result one when added to itself" do not mean any concept with sharp boundaries, i.e. anything that can be used in logic. . . .
>
> But can we not stipulate that the expression "the sum of one object and another object" is to have meaning only when both objects are numbers? In that case, you may well think, the concept *something that gives the result one when added to itself* is one with sharp boundaries; for now we know that no object that is not a number falls under it. E.g. the Moon does not fall under it, since the sum of the Moon and the Moon is not one. This is wrong. On the present view, the sentence "the sum of the Moon and the Moon is one" is neither true nor false; for in either case the words "the sum of the Moon and the Moon" would have to mean something, and this was expressly denied by the suggested stipulation. Our sentence would be comparable, say, to the sentence "Scylla had six dragon necks." This sentence likewise is neither true nor false, but fiction, for the proper name "Scylla" designates nothing.[30]

Frege's argument here is directed against the suggestion that we define the addition sign via a stipulation that terms of the form ". . .+. . ." are to "have meaning only when both objects are numbers." And the objection to this view is that it will leave terms like "the sum of the Moon and the Moon" without reference, and will therefore have the result that the concept *something that gives the result one when added to itself* fails to have sharp boundaries. It is important to note that the failure of what Frege calls the "sharp boundaries" requirement is *not* demonstrated here by pointing to an object that neither determinately falls under nor determinately fails to fall under the concept in question. For under the proposal in question, there is no object denoted by the term "the sum of the

Moon and the Moon." The failure of the sharp-boundaries requirement is instead demonstrated by the fact that the concept-phrase, when saturated by this singular term, gives no truth-value. That is to say, the failure is explicitly the failure of linguistic completeness. The problem Frege points out here is that the proposed stipulation does not provide the concept-phrase "$x + x = 1$" with a reference for each of its saturations by terms of the language. It's a problem that arises only in languages containing singular terms for non-numbers. The criticism here is the same as that directed at Peano above: if the language contains terms for (or quantifiers ranging over) objects that don't satisfy the predicate *F*, then one cannot legitimately restrict the range of arguments for any function referred to in that language to those objects that satisfy *F*.

That the problem is essentially one having to do with linguistic completeness is reinforced as Frege continues the discussion by suggesting and rejecting another way around the difficulty, one that involves a restriction on the formation of concept-phrases:

> If our sentence "the sum of the Moon and the Moon is not one" were a scientific one, then it would assert that the words "the sum of the Moon and the Moon" and the word "one" did not coincide in meaning; but with the stipulation suggested above, the former words would not have any meaning; accordingly we could not truly assert either that their meaning did coincide with the meaning of the word "one" or that it did not coincide with it. Thus it would be impossible to answer the question whether the sum of the Moon and the Moon is one, or whether the Moon falls under the concept *something that gives the result one when added to itself*. In other words, what we have just called a concept would not be a genuine concept at all, since it would lack sharp boundaries. But when once we have introduced the expression "*a* added to *b* gives the result *c*," we can no longer stop the construction of a concept-name like "something that gives the result one when added to itself." If people would actually try to lay down laws that stopped the formation of such concept-names as this, which, though linguistically possible, are inadmissible, they would soon find the task exceedingly difficult and probably impracticable. The only way left open is to give to the words "sum," "addition," etc., if one means to use them at all, such definitions that the concept-names constructed out of the words in a linguistically correct manner stand for concepts with sharp boundaries and are thus admissible.

The "exceedingly difficult and probably impracticable" strategy of restricting the formation of concept-names is a potential way of maintaining linguistic completeness in the face of the problematic stipulation; it is not even in the ballpark of a helpful strategy if the goal is that of ensuring that all functions

are total. And here again, the circumstance described by Frege as one in which the concept in question lacks "sharp boundaries" is one that arises only when the result of completing the concept-phrase with a name from the language has no reference.

Despite the incautious way in which he sometimes puts the requirement of sharp boundaries, a way that when taken in isolation makes it appear that Frege requires e.g. all first-level mathematical functions to be defined on all (*n*-tuples of) objects, mathematical or otherwise, the texts just don't bear this out. As he says in "Function and Concept," an arithmetic that deals just with the integers can quite happily incorporate an addition sign defined over just the integers. What's crucial for Frege is that, within a rigorous theory, no grammatically acceptable concatenation of function-sign and object-sign(s) fails to have reference, which is to say that function-signs must refer to functions that are defined over all of the objects under discussion. If our language has a name for, or quantifiers ranging over, the Moon, then the only acceptable addition sign in that language is one whose reference is defined over the pair <the Moon, the Moon>. But this is the requirement of linguistic completeness.

3.4 Ordinary Discourse

We began with a difficulty apparently posed by the sharp-boundaries requirement for the idea that the sentences of ordinary arithmetic express truth-evaluable thoughts. The difficulty was that if the references of function-terms must be defined on all potential arguments, even those alien to the science in question, then it would follow that the function-terms of ordinary arithmetic lack reference. As we have now seen, this difficulty is not real: Frege's sharp-boundaries requirement is essentially the requirement of linguistic completeness, and is entirely compatible with the definition e.g. of arithmetical functions just over arithmetical objects and with the reference by arithmetical function-terms to just such non-total functions.

But there is a further reason for concern about the status of ordinary sentences in Frege's account: the function-terms of ordinary arithmetical discourse (and ordinary discourse generally) do not in fact meet even the weaker condition of linguistic completeness. The function-terms of *Grundgesetze* do meet it, and it is clear that one can be as careful as Frege is in setting up a scientific language, providing e.g. "dummy" references where necessary, and ensuring by design that such a language meets the requirement. But ordinary arithmetical discourse doesn't meet this standard. It is possible, as Frege points out, to form names to which no reference corresponds: "the least rapidly convergent series" is his own example.[31]

Should we conclude, then, that for Frege only languages meeting the requirement of linguistic completeness can be used to express true and false thoughts?

Here again, the answer is a clear "no." Frege's discussions of the sharp-boundaries requirement, i.e. of the requirement of linguistic completeness, always appear in the context of evaluating languages like that of *Grundgesetze*, languages established for the display of extremely rigorous proofs. It is crucial, given Frege's goal of the syntactic specification of a strictly truth-preserving deductive system, that no well-formed formula lack a truth-value, and hence that the requirement of linguistic completeness is met. But there is no reason for Frege or for Fregeans to insist on such a requirement for languages in general. Specifically, the requirement in question is one whose satisfaction is essential for maintaining maximal proof-theoretic rigor; it plays no role in the use of language for ordinary communication.

In an 1896 letter to Peano, Frege rehearses his reasons for rejecting conditional definitions, emphasizing here that this prohibition, part of the requirement of linguistic completeness, has to do specifically with those contexts in which rigor is at a premium:

> A conditional definition of the sign for a concept decides only for some cases, not for all, whether an object falls under the concept or not; it does not therefore delimit the concept completely and sharply. But logic can only recognize sharply delimited concepts. Only under this presupposition can it set up precise laws. The logical law that there is no third case besides
>
> *a* is *b*
>
> and
>
> *a* is not *b*
>
> is really only another way of expressing our requirement that a concept (*b*) must be sharply delimited. The fallacy known by the name of "Acervus" rests on this, that words like "heap" are treated as if they designated a sharply delimited concept whereas this is not the case. Just as it would be impossible for geometry to set up precise laws if it tried to recognize threads as lines and knots in threads as points, so logic must demand sharp limits of what it will recognize as a concept unless it wants to renounce all precision and certainty. Thus a sign for a concept whose content does not satisfy this requirement is to be regarded as meaningless from the logical point of view. It can be objected that such words are used thousands of times in the language of life. Yes; but our vernacular languages are also not made for conducting proofs. And it is precisely the defects that spring from this that have been my main reason for setting up a conceptual notation.[32]

A similar account appears in 1897, in a general discussion of the requirement of sharp boundaries:

> Inference from two premises very often, if not always, depends on a concept being common to both of them. If a fallacy is to be avoided, not only must the sign for the concept be the same, it must also have the same reference. It must have a reference independent of the context and not first come to acquire one in context, which is no doubt what very often happens in speech.[33]

What "very often happens in speech" is that the concept-word plays its usual role in the expression of thought in ordinary contexts although its behavior in outlying contexts isn't thereby determined. In the case of concept- or function-expressions, the point is that no total concept or function is referred to.

In short, the sharp-boundaries requirement is the requirement not of totality but of linguistic completeness. And it's a requirement for languages intended for maximally-rigorous argument, not for those to be used for ordinary discourse.[34]

3.5 Caesar

Frege's attitude to linguistic completeness and to the scope of function-terms is important to the context of the so-called "Caesar problem" in *Grundlagen*, about which a few remarks are in order. At *Grundlagen* §55, Frege raises the possibility of defining the numbers contextually by saying (1) that 0 belongs to a concept F iff $\forall x{\sim}Fx$; (2) that 1 belongs to a concept F iff $[(\exists x)Fx \,\&\, \forall y \forall z(Fy \,\&\, Fz \rightarrow y = z)]$; and (3) that $(n + 1)$ belongs to F iff $(\exists x)(Fx \,\&\, n$ belongs to $(Fy \,\&\, y \neq x))$. He then explains why such a definition is unsatisfactory:

> We can, of course, by using the last two definitions together, say what is meant by
>
> "the number 1 + 1 belongs to the concept *F*"
>
> and then, using this, give the sense of the expression
>
> "the number 1 + 1 + 1 belongs to the concept *F*"
>
> and so on; but we can never—to take a crude example—decide by means of our definitions whether any concept has the number Julius Caesar belonging to it, or whether that same familiar conqueror of Gaul is a number or not. Moreover, we cannot by the aid of our suggested definitions prove that, if the number *a* belongs to the concept *F* and the number *b* belongs to the same concept, then necessarily $a = b$. Thus we should be unable to justify the expression "*the* number which belongs

> to the concept *F*," and therefore should find it impossible in general to prove a numerical identity, since we should be quite unable to achieve a determinate number. It is only an illusion that we have defined 0 and 1; in reality we have only fixed the sense of the phrases "the number 0 belongs to," "the number 1 belongs to"; but we have no authority to pick out the 0 and 1 here as self-subsistent objects that can be recognized as the same again.[35]

It is easy to read the Caesar passage, taken in isolation, as complaining that the proposed definition fails to assign to e.g. ". . . = 0" a total function, one defined over such objects as Roman generals. Similarly, the complaint would be that the concept-term ". . . is a number" is not assigned a total function. This reading of the passage faces an immediate difficulty when we note that the definitions Frege eventually gives in §§68ff of *Grundlagen* (and of course those he gives in *Grundgesetze* as well) "fail" in exactly the same way: the functions assigned to ". . . = 0" and to ". . . is a number" are defined only over extensions of second-level functions (in the case of *Grundlagen*) or over courses-of-value of first-level functions and truth-values (in the case of *Grundgesetze*). More importantly, as we've seen, Frege doesn't take it to be a general condition of adequacy of function-terms that they are defined over objects outside of the intended domain. So the idea that Frege requires a good definition of these mathematical concept-phrases to assign them functions defined over historical figures is very difficult to square with the texts.

In fact, the proposed *Grundlagen* definitions fail in a considerably more significant way than the suggested failure of totality: it's not that those definitions fail to assign to the function-terms *total* functions; they fail to assign them functions at all. It's crucial for the reconstruction of arithmetical discourse that ". . . = *n*" for each numeral *n* and ". . . is a number" are assigned (first-level) functions, which is to say in part that the individual numerals must be assigned objects. The proposed definitions don't do this; as Frege says, "It is only an illusion that we have defined 0 and 1; in reality we have only fixed the sense of the phrases 'the number 0 belongs to,' 'the number 1 belongs to.'" A straightforward way to dramatize this point is to note that the definitions cannot be employed to give any sense whatsoever to statements of the form "*a* = 0" or "*a* is a number," for *a* any singular term. Frege's choice of "Julius Caesar" for the term *a* doesn't illustrate a point specific to "outlying" singular terms (i.e., to non-arithmetical ones); it's a point that applies to all singular terms. The straightforward solution, and the one that Frege adopts, is to give explicit definitions of the numerals as singular terms and of the term ". . . is a number" as a concept-denoting phrase—but not, of course, as one that denotes a total function.

At *Grundlagen* §106, Frege gives a summary of his discussion of the nature of numbers and of the requirements on definitions of numerical terms, including a

rehearsal of the lesson to be learned from the "Caesar" passage. The relevant part for our purposes is as follows:

> Let us cast a final brief glance back over the course of our enquiry. After establishing that number is neither a collection of things nor a property of such, yet at the same time is not a subjective product of mental processes either, we concluded that a statement of number asserts something objective of a concept. We attempted next to define the individual numbers 0, 1, etc., and the step from one number to the next in the number series. Our first attempt broke down, because we had defined only the predicate which we said was asserted of the concept, but had not given separate definitions of 0 or 1, which are only elements in such predicates. This resulted in our being unable to prove the identity of numbers. It became clear that the number studied by arithmetic must be conceived not as a dependent attribute, but substantivally.

In short, the difficulty with the original proposal has nothing to do with the totality of the concept of number. The difficulty illustrated by the Caesar example is that this proposal fails to treat numbers as objects. In so doing, it also fails to treat the relevant predicates, e.g. ". . . = 0" and ". . . is a number," as function-denoting phrases of any kind.

3.6 Quantification

One final issue, and one about which Frege is simply not very clear, concerns the status of quantified sentences involving function-expressions that do not refer to total functions. Consider the case of first-level concepts. If a concept-expression $g(\xi)$ refers to a concept that is undefined on some arguments (specifically, on arguments with which the science in question is not concerned), then what are the truth-conditions of the universally-quantified sentence $(\forall x)gx$?

Frege notes the connection between the issue of totality and quantification as follows, in a passage that, taken in isolation, would seem to conflict with much of what he has said in the passages quoted above:

> As soon as people aim at generality in propositions they will need in arithmetical formulae not only symbols for definite objects—e.g. the proper name "2"—but also letters that only indicate and do not designate [i.e., bound variables—PB]; and this already leads them, quite unawares, beyond the domain within which they have defined their symbols. One may try to avoid the dangers thus arising by not making the letters indicate objects in general (as I did), but only those of a

> domain with fixed boundaries. Let us suppose for once that the concept *number* has been sharply defined; let it be laid down that italic letters are to indicate only numbers; and let the sign of addition be defined only for numbers. Then in the proposition "$a + b = b + a$" we must mentally add the conditions that a and b are numbers; and these conditions, not being expressed, are easily forgotten. But let us deliberately not forget them for once! By a well-known law of logic, the proposition
>
> "If a is a number and b is a number then $a + b = b + a$"
>
> can be transformed into the proposition
>
> "if $a + b$ is not equal to $b + a$, and a is a number, then b is not a number"
>
> and here it is impossible to maintain the restriction to the domain of numbers.[36]

Frege might be taken here to reject the view that the range of values of a given variable can be restricted to a specific domain, e.g. to numbers. But it is important to see that this is not his point. Frege's argument here is intended to expose the difficulty involved in restricting the definition of function-signs, e.g. of the plus-sign, to a proper part of the collection of objects already under discussion. As he puts it in the previous paragraph:

> Such a restriction would have to be incorporated into the definition, which would thus take some such form as: "If a and b are numbers, then $a + b$ means . . ." We should have a conditional definition. But the sign of addition has not been defined unless every possible complex symbol of the form "$a + b$" has a definite meaning, whatever meaningful proper names may take the places of "a" and "b."[37]

The difficulty is the one already noted above with conditional definitions. We can't coherently restrict the functions referred to by our function-signs to those arguments that satisfy some predicate of the language, since this strategy leads immediately to well-formed combinations of symbols that lack reference. It leaves out not only those singular terms formed by concatenating the function-sign with argument-signs from outside the chosen domain, but also those quantified sentences that involve the function-sign and the negation of the predicate. The point, again, is not that our functions must be total, but rather that they must be defined on all of the arguments over which we quantify in the science in question.

But what, exactly, is the domain of objects over which we quantify in a given science? Frege does not, in explaining the truth-conditions of quantified sentences, refer to an explicitly demarcated range. He explains those truth-conditions variously in terms of the truth of each instance of the matrix and

in more familiar "objectual" ways.[38] In each case, the general idea is that a universally-quantified sentence talks about *all* of the arguments of a certain kind. The unrestricted universality in terms of which Frege explains quantification has often been interpreted as implying, or being, the view that e.g. first-level quantification always talks about *all* objects, second-level about *all* first-level functions, etc.[39] But this can't be right. Recall Frege's claim that as long as we're talking just about e.g. the integers, it's entirely coherent to define the addition sign just over the integers. The sentence "$\forall x \forall y(x + y = y + x)$" is, when the sign is used in this way, true. Hence the range of the first-level quantifier is straightforwardly the range of objects under discussion: in this case, the integers. That Frege is less explicit about the precise limits of the scope of a given domain than his successors have come to be involves a notable lack of comprehensive precision on his part. But given the context of his work, in which, as now, it is entirely commonplace to take one's audience to understand the range of mathematical (or other) objects under discussion, this comparative lack of precision should not strike us as particularly surprising.

It is also worth noting in this context that our own habit now of always explicitly specifying a domain of quantification, rather than relying on a general understanding of the boundaries of the theory in question, is a result of the modern conception of theories as multiply-interpretable sets of sentences whose domains require explicit demarcation on each application. From the Fregean point of view, in which sentences are fully interpreted and theories are sets of thoughts, there is always a determinate subject matter under discussion that already sets the boundaries of quantification. When we're dealing with the arithmetic of the integers, for Frege, "$\forall x$" means "every integer"; when with the reals, "$\forall x$" means "every real," and so on.

That Frege admits the coherence and the scientific acceptability of theories with restricted ranges (e.g., mathematical theories with particular subject-matters) does not by itself tell us much about how he views the ranges of his own quantifiers in *Begriffsschrift* and *Grundgesetze*. Given Frege's view of logic as a purely general science, one might have expected him to view the purely-logical setting of *Begriffsschrift* and *Grundgesetze* as imposing a requirement of complete generality on the ranges of the quantifiers of the languages L_B and L_G. What we have seen is that, at least with respect to L_G, this doesn't seem to be the case. Frege's strategy in *Grundgesetze* is to define L_G's first-level function-terms over just courses-of-value and truth-values. And presumably, then, Frege takes it that the language as a whole, including its quantifiers, is restricted in semantic scope to those objects. (For otherwise, his quantified sentences will all lack truth-values.)

Frege holds that the fundamental principles of logic are purely general: that they hold for all objects and for all functions of appropriate type. The limited range of *Grundgesetze*'s quantifiers brings to light the connection between the fundamental laws of logic and the thoughts expressed by *Grundgesetze*'s axioms.

The latter are restrictions of the former, restrictions to the domain of the system. This it not to say that there is any barrier from the Fregean point of view to the expression of the purely general rules of logic, or to the use of absolutely-unrestricted quantification. But the requirement of rigor for *Grundgesetze*, i.e. the requirement of linguistic completeness, means that the quantifiers of L_G can only range over that domain over which its function-terms are defined.

3.7 Conclusion

Despite some often-quoted passages that seem to indicate that Frege takes all functions to be total, there is in the end no coherent way to read him as endorsing this requirement. In the notorious passages in question, Frege insists that argumentative rigor demands (1) the rejection of conditional and piecemeal definitions, and (2) the satisfaction of the requirement of linguistic completeness. Hence in any context in which argumentative rigor is at a premium, it is important that every function be defined over every argument of the domain under discussion. That even the strictest demands of rigor fail to imply a requirement of totality on the part of functions is made most clear by the fact that the function-expressions of Frege's own extraordinarily careful and rigorous language of *Grundgesetze* fail to refer to total functions.

Furthermore, the fact that the function-expressions of ordinary language and of ordinary scientific discourse are not defined over outlying objects and concepts—e.g. the fact that ". . .+. . ." as ordinarily used is not defined over mountains or proper sets—tells us nothing, from the Fregean point of view, about the capacity of those expressions to appear in sentences expressing truth-evaluable thoughts. And finally, the fact that the function-expressions of ordinary discourse fail to have sharp boundaries even within the intended domain of discourse—e.g. that "heap" is vague, that "the eldest son of . . ." often delivers no value when attached to the name of a person—tells us just that these expressions are not ideally suited to the rigors of argument, not that they cannot be used, according to Frege's semantic theory, in the expression of truths and falsehoods.

Notes

1. For a fuller treatment of this issue, see Blanchette [forthcoming].
2. *Grundgesetze* (hereafter *Gg*) 2 §56: "a concept that is not sharply defined is wrongly termed a concept." See also the §62 discussion of "inadmissible sham concepts."
3. [1897c] p 168/155.
4. *Gg* 2 §63.

5. *Gg* 2 §65.
6. See *Gg* 1 §31.
7. [1891] p 20/148.
8. See, e.g., *Gg* 1 §§5,8; *Gg* 2 §56; [1906c] p 195/180.
9. See, e.g., [1892a] and [1918], in which Frege treats sentences about historical figures and mundane objects as expressing truth-evaluable thoughts.
10. [1892c] p 196/185 note.
11. Weiner [2007] p 702.
12. Weiner [2007] p 703.
13. [1891], [1892a], [1892c], [1918a], [1918b], [1923].
14. Or, as one should more carefully put it, his *attempt* to remedy this defect.
15. *Gg* 1 §10. I have substituted modern universal-quantification notation for Frege's own.
16. Ibid.
17. Ibid.
18. Ibid. This passage may appear to be in conflict with Frege's prohibition against piecemeal definitions, discussed below. But presumably his point is that in a context in which we will refer to new such functions or objects, we will build a new system from the ground up, one whose function-terms refer to functions defined over all of the new domain. See §3.3.
19. *Gg* 1 §32.
20. [1891] p 19/148.
21. See, e.g., the letter to Peano of September 29, 1896, especially [1980] pp 114–115; [1897c] p 168/155.
22. *Gg* 1 §31.
23. [1891] p 19/148.
24. [1897c] p 168/155.
25. Ibid.
26. *Gg* 2 §§56, 57.
27. *Gg* 2 §58.
28. *Gg* 2 §60.
29. Reck and Awodey [2004] p 155, from the summer semester 1914.
30. *Gg* 2 §§63–64.
31. [1892a] p 28/159.
32. Letter to Peano of September 29, 1896; [1980] pp 114–115.
33. [1897c] p 168/155 (translation slightly amended).
34. See also *Gg* 2 §57, in which the sharp-boundaries requirement is explicitly restricted to the "exact sciences."
35. *Grundlagen* (hereafter *Gl*) §56.
36. *Gg* 2 §65.
37. Ibid.
38. See *Begriffsschrift* §11; *Gg* 1 §§8, 25, 31.
39. See, e.g., Goldfarb [2001] p 27: "Quantifiers in Frege's system have fixed meaning: they range over all items of the appropriate logical type (objects, one place functions of objects, two place functions of objects, etc.)."

4

The Analysis of Arithmetic

4.1 The Issue

To return to the central issue: Frege claims to be engaged in a demonstration that the truths of arithmetic are purely logical, a demonstration that proceeds via proofs of arithmetical truths from purely-logical ones. But the proofs he gives do not in any obvious sense terminate with truths of arithmetic. They terminate instead with thoughts about such unfamiliar things as courses-of-values and highly convoluted functions defined on them. As we have seen, Frege represents the proven truths as the results of analyses of arithmetical truths: the central idea is that those arithmetical truths are clarified or elucidated via an analysis of them and of their components, and only subsequently are the thus-analyzed truths subjected to proof. As Frege puts it in 1914:

> The effect of the logical analysis . . . will then be precisely this—to articulate the sense clearly. Work of this kind is very useful; it does not, however, form part of the construction of the system, but must take place beforehand. Before the work of construction is begun, the building stones have to be carefully prepared. . . .[1]

The preparation of the "building stones" is the analysis of arithmetical truths, an analysis that yields those fully-articulated truths that will then be subject to proof. But the rather large gulf between the analysandum-thoughts and their analysans-counterparts leaves it far from clear that what Frege gives can in any reasonable sense be called merely an "analysis" of the ordinary contents of arithmetical discourse. More to the point: because of Frege's lack of clarity about the nature of what he calls "analysis," it is not obvious that the proof from purely logical principles of his highly-complex and unfamiliar analysans-thoughts can really be taken as a guarantee of the purely logical nature of the truths of arithmetic.

The purpose of this chapter is to clarify, insofar as this is possible, Frege's understanding of the role of conceptual analysis in his proof-theoretic project.

Frege's analyses themselves were failures: his account of arithmetic rests on the premise that there are Fregean extensions, and we now know that there are no such things. This result marks the end of the specific Fregean analytic strategy: there is no way to do quite what Frege had hoped to do, i.e. to cash out the content of arithmetic as a body of truths about arithmetically-behaved extensions. But there is still a central and important question at the heart of his project that rewards investigation. That is the question of how Frege understood the task of conceptual analysis, and of whether he was right that analysis of this general kind can be informative about the logical properties and relations of thoughts so analyzed. One wants to know, in short, whether a proof of a highly-analyzed version of a given thought τ from premises P can establish that τ itself follows logically from P when the analysis in question is of a Fregean kind.

Frege nowhere says what an adequate analysis is. His strategy is not to present a theory of analysis but instead to present specific analyses that he assumes his audience will recognize as successful. His approach is in this way just that of the analytic projects he himself cites as paradigms: as he says, the development of mathematics has depended crucially on the "sharpening" of the "concepts of function, of continuity, of limit, and of infinity . . ." and of those of "negative and irrational numbers."[2] Frege cites this important clarificatory work, none of which, of course, comes along with an account of the success-conditions of analysis, as precedent for his own. "Proceeding along these lines," i.e. along the lines of analysis that have given us the precise modern accounts of infinity and of continuity, Frege says, "we are bound eventually to come to the concept of number," i.e., to the analytic project of *Grundlagen*.[3]

Some of Frege's analyses seem clearly in the spirit of those he cites, namely, analyses whose adequacy, though hard to characterize precisely, is straightforwardly recognizable once the analysans is put to work in the context of whole judgments and lines of reasoning. His *Begriffsschrift* account of the ancestral of a relation, his 1881 account of the relation *n is a multiple of m*, and his 1884 account of the relation of equinumerosity are all straightforwardly recognizable as doing what good conceptual analyses do: they offer useful clarification while in some hard-to-specify way managing to stick closely enough to the original subject-matter that they are not reasonably charged with changing the subject.

Once the analyses become more sophisticated in Frege's project, and particularly by the time they reach the all-important account of cardinal number, things are less clear. The distance between analysans and analysandum is relatively large, and it is not obvious, for reasons discussed below, that Frege's intention is to provide analyses that fit the general framework of those he cites as paradigms. Part of the difficulty here is simply the venerable paradox of analysis: because only an entirely trivial restatement of an analysandum is immune from the potential charge of failing to hit its mark, and no such trivial restatement can count as providing an elucidation at all, the goal of providing accurate yet informative analyses

is guaranteed to be elusive. For the same reason, the demand that one provide a fully-general but precise account of the conditions for accurate yet informative analysis is too much to ask. Nevertheless, just as it is sometimes clear that an analysis has succeeded, even in the absence of clearly-articulated success-criteria, it's also occasionally clear that a proffered analysis has in fact strayed too far from its original subject-matter to count as successful. The central questions to be asked in what follows are those of (1) roughly what kind of relationship Frege understood to obtain between the analysandum-thoughts of ordinary arithmetic and their analysans-counterparts, and (2) whether Fregean analyses, so understood, overstep the kind of boundary just indicated, leaving Frege's proofs to stand merely as demonstrations of the purely-logical grounds of something other than arithmetic.

4.2 Analysis as Thought-Preserving?

Some of Frege's texts suggest that the relation between analysandum- and analysans-thought is supposed to be that of identity. First are those passages in which Frege refers to specific proven thoughts via the vocabulary of ordinary arithmetic and claims that his proofs are proofs of arithmetical truths.[4] Second are those passages in which Frege seems to talk explicitly about analysis as thought-preserving. Recall, e.g., the passage from [1914] quoted in §1.1 above, in which Frege notes:

> [I]t may not be possible to prove a truth containing a complex constituent so long as that constituent remains unanalyzed; but it may be possible, given an analysis, to prove it from truths in which the elements of the analysis occur.[5]

Presumably the thing referred to as "it" here is the thought in question, the very thing that is unprovable prior to analysis and provable thereafter, an entity that survives the analysis. To repeat a passage just quoted (§4.1):

> The effect of the logical analysis . . . will then be precisely this—to articulate the sense clearly.[6]

These passages echo the idea presented in *Grundlagen* in terms of contents, that a given content can sometimes be proven only after an analysis of its constituents, implying, one presumes, that the content itself survives that analysis.

But there are some straightforward difficulties with the idea that content-preservation, and especially, later, thought-preservation, is a condition of adequate analysis in Frege's work.

The first difficulty is that the idea of analysis as thought-preserving is incompatible with what is perhaps the most straightforward account of thought-identity extractable from Frege's work, the account in terms of what we have called above the "cognitive criterion" of thought-individuation. According to this criterion, two sentences express the same thought only if they are, roughly, fairly readily-recognizable synonyms: specifically, only if anybody who understands both sentences and affirms one of them must, on pain of incoherence, affirm the other. This way of understanding thoughts would seem to indicate a relatively fine-grained criterion of individuation, and one that does not sit easily with the idea that the two sentences involved in a non-trivial analysis express the same thought. It is, one might think, not at all incoherent for a person who understands both of

(β) One is the successor of zero

and

(β*) For some concept *F* and some object a falling under *F*, the extension of the concept *equinumerous with the concept identical with the extension of the concept (equinumerous with the concept (not self-identical))* is the extension of the concept *equinumerous with the concept F*, and the extension of the concept *equinumerous with the concept not self-identical* is the extension of the concept *equinumerous with the concept falling under F, but not identical with a*

to affirm the first without affirming the second. Similarly for many of Frege's important analyses. In short, if thoughts are individuated in the fine-grained way just suggested, then Frege's analyses, and indeed all non-trivial analyses, must fail to be thought-preserving.

One might try to defend the view of Fregean analyses as thought-preserving in the face of an imagined audience's differing attitudes toward each of a pair of analysandum- and analysans-sentences by holding that the audience in such cases fails to understand at least one of the sentences. Frege is clearly at least tempted by this "way out" in [1914]. Answering the worry in question as applied to that portion of a thought expressed by a simple sign of the ordinary sentence, Frege says:

> Perhaps there still remains a certain unclarity. How is it possible, one may ask, that it should be doubtful whether a simple sign has the same sense as a complex expression if we know not only the sense of the simple sign, but can recognize the sense of the complex one from the way it is put together? The fact is that if we really do have a clear grasp

> of the sense of the simple sign, then it cannot be doubtful whether it agrees with the sense of the complex expression. If this is open to question although we can clearly recognize the sense of the complex expression from the way it is put together, then the reason must lie in the fact that we do not have a clear grasp of the sense of the simple sign, but that its outlines are confused as if we saw it through a mist. The effect of the logical analysis of which we spoke will then be precisely this—to articulate the sense clearly.[7]

This way of talking, in which a later clear grasp of a previously only inchoately-understood content or thought involves what Frege describes metaphorically as a clearing-away of an obscuring cloud, often appears in Frege's work when he talks about familiar cases of conceptual refinement in the historical development of mathematics. And here Frege clearly has a point: part of what is gained as mathematical disciplines develop is greater clarity about their fundamental components. As he puts it in *Grundlagen*:

> Often it is only after immense intellectual effort, which may have continued over centuries, that humanity at last succeeds in achieving knowledge of a concept in its pure form, in stripping off the irrelevant accretions which veil it from the eyes of the mind.[8]

If we read such conceptual refinements as involving something like the increasingly-clear grasp of a fixed collection of thoughts, one might take this phenomenon as a model of what goes on with Fregean analyses: the analysandum-sentence and analysans-sentence express the same thought, and the fact that the ordinary speaker does not recognize the sentences as synonyms is a symptom of that speaker's failure to grasp the first thought with sufficient clarity.

But there are serious problems with this attempt to maintain simultaneously the theses (1) that analysis preserves thoughts, (2) that analysis can be highly non-trivial, and (3) that thoughts are individuated in the way indicated by the cognitive criterion. There is, first, the implausibility of the idea that ordinary speakers fail to understand what is meant by such simple sentences as "3 is greater than 2." Secondly, this picture makes it entirely mysterious how the sentences of ordinary arithmetic could have come to express the senses they do, if those senses had been, prior to Frege's work, grasped by nobody.[9] Finally and most importantly for our purposes: if the criterion of successful analysis is thought-identity, but the identity of the thoughts expressed by analysandum- and analysans-sentences is not something that can straightforwardly be assessed by those who have an ordinary grasp of what those sentences say (when this falls short of the rarefied level of "understanding" suggested by Frege above), then the criterion is of no use in assessing the correctness of a given

analysis. To know that thought-identity is the mark of successful analysis is, under these conditions, to know nothing that will help distinguish successful from unsuccessful analyses.

A different and more plausible way of maintaining the idea of analysis as thought-preserving in the face of the highly non-trivial nature of analyses is the idea, also pursued at various points by Frege, that thoughts are individuated in a more coarse-grained way than that suggested by the cognitive criterion. As we've seen, Frege is on occasion sympathetic to the idea that two sentences express the same thought if, logical truths (and falsehoods?) aside, those sentences express logical equivalents. But here too there are difficulties. In addition to general difficulties of circularity that face such a criterion, the central difficulty for present purposes is that the criterion is inapplicable in the relevant cases. Before knowing of a proposed analysans-sentence s^* whether what it expresses is logically equivalent with what's expressed by the analysandum-sentence s, we will typically need already to know whether the analysis is an accurate one.[10] We need, for example, to know whether claims about numbers can in fact accurately be understood via an analysis in terms of extensions before we can answer the question of whether a given thought about numbers is logically equivalent with a corresponding one concerning extensions.

In short, there are texts in which Frege seems to take thought-preservation to be the goal of conceptual analysis. But these texts do not get us very far toward a clear understanding of Frege's conception of analysis, for two reasons. First of all, there are obvious tensions between these texts and other parts of the Fregean corpus, with the result that it is not clear to what extent these remarks can be taken to be an indication of his considered opinion. Secondly, and more importantly for our purposes, the criterion itself is virtually empty, given Frege's lack of clarity about the conditions under which two sentences express the same thought. Without a clear account of those conditions, an account that Frege appears never to have felt the need to provide, the idea that an analysans-sentence expresses an adequate analysis of what's expressed by an analysandum-sentence just when those sentences express the same thought is of little help in pinning down the Fregean criterion of adequate analysis.

4.3 Reference-Preservation and Analysis

One might have thought that an adequate analysis of arithmetic ought to issue in reference-preserving terms and phrases, so that, e.g., a sentence expressing a proposed analysis of the thought expressed by "4 > 2" would count as sufficiently conservative over its original only if it has (possibly complex) singular terms referring to 4 and to 2, and a (possibly complex) relation-term referring to the

greater-than relation.[11] But as has been widely discussed, Frege clearly does not aim for such term-by-term reference-preservation.[12] And it might be thought that at this point Frege's intended analyses have decisively crossed that hazy boundary-line separating successful analysis from unsuccessful, i.e., from mere change of subject-matter. Or, to put the point in terms of textual exegesis, it might be thought that Frege's clear lack of interest in preserving the reference of individual arithmetical terms indicates that he was not, after all, involved in a project reasonably described as that of conceptual analysis.

The clearest indication that Frege did not intend term-by-term reference preservation is that he takes there to be a certain amount of arbitrariness involved in the assignment of reference to the numerical terms of the languages of *Grundlagen* and *Grundgesetze*. Frege explains as follows the choice of specific extensions in *Grundlagen* as a way of circumventing a "difficulty" involving the determinacy of numerical reference:

> This way of getting over the difficulty cannot be expected to meet with universal approval, and many will prefer other methods of removing the doubt in question. I attach no decisive importance to bringing in the extensions of concepts at all.[13]

That is to say: though the numeral "0" is taken in *Grundlagen* to refer to a particular extension, there are alternative, equally-good ways of settling its reference—even, it seems, ones that take that numeral to refer to something that isn't an extension. Frege later revises his view about the dispensability of extensions; as becomes clear by the time of *Grundgesetze*, the only objects that can play the role of numbers are extensions, or, more generally, the courses-of-values of functions.[14] But what seems relatively clear from this passage is that Frege does not see himself as first determining exactly which objects the ordinary numerals refer to and then making sure that the numerals of his newly-fashioned sentences refer to just those objects.

Further evidence that Frege is not interested in reference-preservation as applied to numerals is the difference between the *Grundlagen* and *Grundgesetze* analyses: in the earlier account, numerals refer to the extensions of second-level functions, while the later, technically neater account has them referring to the extensions (courses-of-values) of first-level functions. Frege gives no indication of having changed his mind about the "right" reference of the numerals, and seems to view the change simply as one of technical convenience. Once again, Frege's procedure can't coherently be read as that of first determining to which objects the ordinary numerals refer, and then ensuring that the numerals of the analysans-sentences refer to exactly those things. Further: given the interdefinability of function- and singular terms, the same holds for the predicative phrases of ordinary arithmetic. Because Frege is not concerned with preserving

the reference of e.g. "2," he cannot be said to be concerned with preserving the reference of "cardinal number" or of "greater-than" either.

Paul Benacerraf has argued that this shows Frege not to have been engaged in the kind of analytic project described here, but to have been interested in providing something more like a model of arithmetic.[15] In a similar vein, Joan Weiner argues that the failure to preserve reference helps to establish that Frege was interested not in providing an analysis of arithmetic—i.e., of the content of ordinary arithmetical discourse (which, recall, Weiner takes Frege to view as semantically defective)—but instead in providing a replacement for that discourse.[16] On Weiner's view, the truths whose logical grounding was to have been demonstrated via the proofs in *Grundgesetze* are those of a new replacement science, not of one that dates back to Euclid.

Though there is no crystalline division between "analysis" and "replacement," especially within mathematics, we can see just how far Frege's project is from an analytic one, on Weiner's reading, by noting what in Weiner's view Frege's replacement for arithmetic is supposed to preserve from its original. As Weiner sees it:

> Faithful definitions must be definitions on which those sentences that we take to express truths of arithmetic come out true and on which those series of sentences that we take to express correct inferences turn out to be enthymematic versions of gapless proofs in the logical system.[17]

In short, the project is required simply to preserve the (apparent) truth-values and derivability-relations of the sentences of arithmetic as ordinarily understood.[18] This account is fairly-naturally motivated by Frege's lack of interest in preserving reference. After all, one might argue, if the references of the arithmetical terms aren't preserved, then the semantic similarity between ordinary and Fregean sentence is very thin indeed, too thin for the latter to count as expressing an analysis of what's expressed by the former. What we're left with is a new, improved science, one whose deductive relations mirror (for the most part) those of the original but inadequate science.

We'll see in §4.5 that the inference just noted—from the failure of Frege's analyses to preserve reference to their failure as analyses—is unwarranted. But first it is important to see that the picture of Frege as providing simply a model of or a replacement (in the sense just given) for arithmetic does not cohere well with Frege's general views about the point and the implications of his logicist reduction.

If a *reduction* of a theory T to a theory T' requires simply the preservation of (apparent) truth-values and derivability-relations, then reductions are relatively easy to come by. If for example we can construct in T' an isomorphic copy of the

subject-matter of T, then we can use the resources of T' to define T's terminology in a way that preserves truth-values and derivability-relations, and furthermore allows us to derive the theorem-sentences of T, so understood, from T'. Consider, for example, the empiricist account of arithmetic as caricatured by Frege.[19] Starting with some mundane facts about pebbles—e.g., that a pebble can always be moved 1 inch to the right—we can construct an ω-sequence of pebble-positions. It's now a simple matter to use pebble-positions to define arithmetical terminology, and to prove the resulting contents of, e.g., "every natural number has a successor" by appealing to pebble-axioms. We can easily preserve (apparent) truth-values and derivability relations under this interpretation.

In the other direction: any axiomatized theory at all whose domain is isomorphic to some construction out of Fregean extensions is "reducible" in this way to Frege's logic. Reductions of this kind are useful in clarifying the structure of a theory and in giving relative-consistency proofs. But they are relatively uninformative about the nature or the grounds of the reduced theory. A reduction, in the sense in question here, is essentially a reinterpretation of the language of T according to which, as newly understood, its sentences express claims about the subject-matter of T'. Hence the question of reducibility in this sense has little to do with whether the subject-matter originally treated by T has anything in common with the subject-matter of the replacing theory.

Frege took himself to be engaged in a project that, if successful, would have demonstrated that arithmetic is grounded in pure logic, in a way in which neither geometry nor any empirical science is so grounded. The fundamental truths of geometry, says Frege, are synthetic, since they are the kinds of truths that can in principle be denied without self-contradiction. The possibility of coherent denial forms no barrier to a "replacement"-style reduction of geometry to a body of analytic truths: it is unproblematic to reinterpret the terms "point," "line," etc. in such a way that Euclid's axiom-sentences express analytic truths about collections of extensions and relations thereon. But the synthetic nature of the geometric thoughts does stand firmly in the way of a Fregean reduction of geometry to anything analytic: no analysis of the thoughts actually expressed by Euclid's axiom-sentences can show those thoughts to be grounded in pure logic.[20]

Frege takes it that his account of arithmetic gets the subject-matter right, in a way that reductions meeting just the replacement-criterion can't be relied upon to do. His extended arguments to the effect that arithmetic is not about ideas, that it is not about empirically-given objects or properties, and that its truths are not revealed by intuition are all part of his project of correctly analyzing arithmetical contents. As Frege says in criticism of the empirical approach, "Mill understands the symbol + in such a way that it will serve to express the relation between the parts of a physical body or of a heap and the whole body or heap; but such is not the sense of that symbol"; and that "In order to be able to call arithmetical truths laws of nature, Mill attributes to them a sense which

they do not bear."[21] In response to Hankel's attempt to define addition in terms of our intuitions of magnitude, Frege objects:

> The definition can perhaps be constructed, but it will not do as a substitute for the original propositions; for in seeking to apply it the question would always arise: Are Numbers magnitudes, and is what we ordinarily call addition of Numbers addition in the sense of this definition?[22]

The response to Newton contains a similar worry. Even if Newton's definitions can be given in a noncircular way, Frege claims, "Even so, we should still remain in doubt as to how the number defined geometrically in this way is related to the number of ordinary life. . . ."[23]

Similar considerations make it clear that Frege was not interested simply in giving a model of arithmetic. That there is a model of arithmetic in the domain of Fregean extensions (or that there would have been if those extensions hadn't been problematic) is itself an interesting mathematical fact, and one well worth pursuing independently of the reasons Frege had for its pursuit. But Frege was interested in demonstrating that arithmetic is grounded in logic; and the fact that there is a *model* of arithmetic in the domain of logic does nothing to demonstrate this, any more than the existence of a model of geometry in arithmetic demonstrates that geometry is grounded in arithmetic. In this vein, it's worth keeping in mind Frege's description at the beginning of *Grundgesetze*:

> Mathematicians generally are indeed only concerned with the content of a proposition and with the fact that it is to be proved. What is new in this book is *not the content of the proposition*, but the way in which the proof is carried out and the foundations on which it rests.[24]

Whatever Frege was trying to preserve in his analyses, it wasn't just structural similarity of domain, and it wasn't just the (apparent) truth-values and derivability-relations among sentences: these criteria are far too weak to make sense of Frege's central claims about the nature of arithmetic and about the implications of his *Grundgesetze* proofs.

To summarize, Frege clearly does not intend his analysis of arithmetical truths to culminate in analysans-sentences whose singular and function-terms refer to the same objects and functions as do their ordinary counterparts. To put the point slightly more carefully: he does not take it that each of the singular and function-terms of ordinary arithmetic has a unique content or reference that any successful analysis of the science must preserve as the contents or references of the corresponding terms of the fully-analyzed version of the science. This might lead one to think that Frege's reductionist project consists simply of the provision of a model, or of a replacement science that preserves apparent

truth-values and derivability-relations. But this account makes no sense of what Frege took to be the whole point of his logicist reduction, which was to show that arithmetic is grounded in logic in a way that is epistemologically and ontologically significant: he took it that arithmetical truths are themselves truths of logic, and that his reduction would demonstrate this. The question is that of whether one can square this strong sense of reduction, which requires a considerably stronger semantic relationship between ordinary and Fregean arithmetical sentences than would be required by the modeling or replacement strategies, with the failure of Frege's reductions to preserve reference.

The claim pursued in Section 4.5 is that the answer here is "yes"—that an analysis compatible with the kind of arbitrariness under discussion here can suffice for Frege's purposes. We'll see this by looking in some detail at *Grundlagen*'s analyses and Frege's commentary on them. But first we pause to examine another response to the issue of arbitrariness.

4.4 Dummett on What's Preserved

The central question at issue is this: how faithful to the original content or sense of the analyzed expressions do Frege's analysantia have to be in order for them to play their intended role within his project? How close in content or sense, for example, do Frege's defined replacements for "cardinal number" and "successor" have to be to their originals in order for Frege's derivation of the relevant sentence to stand as a proof of the thought that every cardinal number has a successor?

Michael Dummett's answer to the central question, attending particularly to the difficulty posed by the arbitrariness of Frege's definitions, is twofold: (1) that the sense of the newly-defined term must come as close as possible to capturing the ordinary sense; and (2) that, when the sense of the newly-defined terms goes beyond that of the original, incorporating an arbitrary element to close the gaps left by the original sense, the important requirement is that the arbitrary part play no essential role in the derivation of the sentence in question. The arbitrary part of the definition must be, as it were, quarantined from the proof-theoretic part of the project. As Dummett explains the second requirement:

> If the proof of the statement does not depend essentially on the features of the definition that have been arbitrarily chosen, it will turn only on those features that were responsive to the sense of the expression as it is ordinarily used; and then the statement, as ordinarily understood, will have been adequately demonstrated.[25]

One difficulty with this reading is that it does not fit well with Frege's extreme care regarding derivations, and particularly regarding his emphasis on ensuring

that no step can "slip in unnoticed" in his derivations, as a guarantee of the purity of the reasoning expressed.[26] Dummett's approach would have it that a crucial feature of Frege's derivations and the proofs they express is that they make no appeal to the arbitrary parts of definitions. The central idea here with respect to the cardinal numbers is that the proofs appeal only to what all of the acceptable numerical reference-assignments have in common, namely, their implication of Hume's Principle, and make no appeal to any of the idiosyncratic features of the chosen assignment. But although, as Richard Heck has argued, the derivations in Frege's treatment of cardinal arithmetic make no *essential* appeal to the definition of the numbers as extensions beyond the fact that this definition allows one to derive Hume's Principle, the qualifier "essential" is important here.[27] Frege's derivations employ the course-of-values notation frequently and importantly, and they appeal to the fact that the number-terms refer to particular courses-of-values. The fact that these appeals can be eliminated in principle is by no means obvious, and is nowhere insisted upon or even remarked by Frege. If it had been Frege's intention to guarantee the purity of his proofs by making sure that those proofs made no appeal to the arbitrary "part" of his definitions, he would presumably have provided his audience with some way of distinguishing such appeals from other, legitimate, steps of derivation. But he does no such thing.

More importantly, the idea of avoiding the problem posed by the arbitrariness of Frege's definitions via the proof-theoretic isolation of their arbitrary component mis-locates the difficulty itself. The worry one might hope to address by such proof-theoretic isolation can be characterized roughly as follows: one worries that the incorporation of "arbitrary" features into the reconstructed versions of arithmetical truths opens up the possibility of contaminating the proofs in which they occur, bringing into those proofs an illegitimate appeal to some truth based on the arbitrary, i.e., non-arithmetical, content of those truths. The requirement that the arbitrary part of the definitions play no role in those proofs is intended to head off this potential source of contamination.

But the potential problem posed by the arbitrariness of Frege's definitions is not an issue of proof-purity. Frege's proofs are all "clean": each step in a proof is either a logical axiom, a consequence of previous steps via an acknowledged rule of logical inference, or the unpacking of a stipulative definition. Each appeal to a definition is, as above, entirely dispensable: every formula Frege derives expresses a thought that is provable via appeal only to laws of logic. Hence every proven thought is (leaving aside the issue of the inconsistency of Frege's system) unquestionably grounded in pure logic. No claim regarding the inclusion or lack thereof of appeals to "arbitrary" features of definitions is relevant here. The potential difficulty posed by the arbitrary nature of some of Frege's definitions stems from the undoubted fact that this arbitrariness undermines the idea that Frege's definitions preserve the reference of individual terms. That the definitions fail to preserve term-reference might be taken to imply that the thoughts

expressed by the sentences in question, and particularly by the derived sentences, are not sufficiently similar to the truths of arithmetic to serve Frege's purposes. If in fact the proven thoughts are not sufficiently similar to the truths of arithmetic, then Frege's project fails; if they are sufficiently similar, then (again ignoring the inconsistency of the system) it succeeds. The question raised by the arbitrariness of Frege's definitions has entirely to do with the content of the derived formulas, not with the principles employed in their derivations.

4.5 What's Preserved

To recap: the singular terms and function-terms, including the concept-phrases, of Frege's analysans-sentences do not, and were clearly not intended to, share reference with those parts of the ordinary arithmetical sentences to which they correspond. The worry is that this large semantic gap between the sentences of ordinary arithmetic and Frege's derived sentences undermines the idea that the purely-logical grounding of his proven truths is a guarantee of the purely-logical grounding of the truths of arithmetic.

But this initial appearance of difficulty is misleading. Here we look at the details of Frege's *Grundlagen* reduction of the finite cardinals, by way of clarifying what he took to be the purpose of that analysis.[28] We'll see that, given Frege's understanding of the connection between sentential contents or senses and the references of constituent terms, the failure to preserve reference term-for-term is no barrier to the preservation of what matters, at the level of sentences, for Frege's project.

4.5.1 The Case of Directions

Frege holds in *Grundlagen*, as he has held at least since *Begriffsschrift*, that the content of an entire sentence is at least sometimes "prior" to the content of at least some of its constituent terms, in the sense that we arrive at a clear appreciation of the content of those terms by first understanding whole judgments in whose expression they figure, and then parceling out the content to those constituents. For illustration, we return to the case of directions and parallel lines already discussed (Section 1.3).[29] According to Frege, we come to understand the term-forming operator "the direction of . . ." by first understanding whole sentences of the form:

(dir) The direction of line a = the direction of line b

which we do by noting the equivalence between these sentences and the corresponding sentences of the form

(par) Line *a* is parallel to line *b*

As Frege says about this case:

> The concept of direction is only discovered at all as a result of a process of intellectual activity which takes its start from the intuition.[30]

Intuition gives us lines and the relation of *parallelness*. A further "process of intellectual activity" is required to get us from there to the concept of *direction*. And as Frege portrays it, what we begin with, and what we obtain by the intellectual re-carving, are whole judgment-contents. Having realized that *parallel* is an equivalence-relation, we note that (par) "can be taken as" an identity: we re-carve (par) as (dir). It's our understanding of the connection between the contents expressed by the two kinds of sentences, i.e., of the fact that sentences of the form (dir) express just what their counterparts of the form (par) do—that counts as an understanding of *direction*. To know what directions are is to just to know that they're what parallel lines have in common.

Immediately following the discussion of directions and parallel lines, Frege notes that we can (or perhaps that we do) engage in similar re-carving strategies with respect to e.g. claims about parallel planes, to obtain the concept of *orientation*, with respect to claims of geometrical similarity to obtain the concept *shape*, and so on.[31] The fundamental idea throughout is the same: some objects—e.g., directions, orientations, and shapes—are understood via an understanding of their identity-conditions, which themselves are understood via a prior understanding of equivalence-relations that obtain between already-understood entities.

Frege next asks how we might define (dir). The question is that of how we might give an account of (dir)'s content in a way that's more fully-articulated than is (dir) itself. (Never mind, for the present discussion, that it's hard to see what point there would be in doing so; the example is purely illustrative for Frege.) One's initial thought, as Frege points out, might be simply to define (dir) via (par). But as natural as this might seem, it won't do for Fregean purposes: Frege needs (or will need, when we turn to the "real" case of numbers) an analysis that maintains the identity-statement structure of (dir), i.e., one that proceeds by giving an account of the individual singular terms themselves, the terms of the form "the direction of. . . ." Frege's elegant solution is as follows:

> If line *a* is parallel to line *b*, then the extension of the concept 'line parallel to line *a*' is identical with the extension of the concept 'line parallel to line *b*'; and conversely, if the extensions of the two concepts just named are identical, then *a* is parallel to *b*. Let us try, therefore, the following type of definition:
>
> the direction of line *a* is the extension of the concept 'parallel to line *a*'.[32]

That is, we give an account of (dir)'s content via an account of the content of each of its singular terms. Defining the direction-terms in this way, (dir) becomes short-hand for:

(dir*) the extension of the concept *line parallel to line a* = the extension of the concept *line parallel to line b*.

The important point about (dir*) is the strong equivalence its content bears to the ordinary content of (dir), if Frege is right about the two steps that get us from one to the other. The first is the Fregean claim that the ordinary content of (dir) is equivalent to that of (par)—that, as one might put it, (par) offers a reasonable analysis of the ordinary content of (dir). This is a claim for which there can be no decisive proof; it's something one can only see, as Frege puts it, "by an immediate insight."[33] The second step is that from (par) to (dir*), a step of straightforward logical proof. That proof requires a principle that Frege assumes in *Grundlagen* and that we might call the "Principle of Extensionality," (PE).[34] This is the principle that sentences of the form

(EXT) The extension of F = the extension of G

are logically equivalent with their counterparts of the form

(UG) $\forall x(Fx \text{ iff } Gx)$

Taking (PE) as a simple and fundamental principle of logic, it's an easy and straightforward matter to prove (dir*) from (par) and vice-versa.[35]

If Frege had been right (a) that (par) offers a good analysis of the ordinary content of (dir), and (b) that (dir*) is a demonstrable logical equivalent of (par), then: any proof from premises P of an instance of (dir*) would straightforwardly show that P suffices to ground the corresponding instance of (dir), as ordinarily understood. For any premises sufficient to ground an instance of (dir*) must suffice to ground the corresponding instance of (par) by the logical equivalence (b), and therefore to ground the corresponding instance of (dir), by the adequacy of the analysis (a). In short, the "transfer" of grounding would in this case be due to the fact that what (dir*) expresses is logically equivalent to a good analysis of the ordinary content of (dir).

Frege's definition of the term-forming operator "the direction of . . ." preserves what is important with respect to sentences of the form (dir). But this is not to say that the definition preserves the reference of the ordinary direction-terms. The definition Frege gives is no better (except aesthetically) than are a whole host of alternative definitions that assign distinct objects to the

direction-terms. Consider e.g. the definition on which "the direction of line *x*" is taken to stand for the extension of the second-level concept within which fall all and only those first-level concepts under which fall all and only those lines parallel to *x*. (To speak set-theoretically for a moment: while Frege's actual definition takes "the direction of line *x*" to stand for (roughly) the set of lines parallel to *x*, the alternative just noted takes the term to stand for the singleton of that set.) This alternative definition does just as well as the original: the sentence (dir), understood in terms of the new alternative definition, expresses a content that is demonstrably logically equivalent with the content ordinarily expressed by (par). Any premises that logically entail the one content will suffice to entail the other, and vice-versa. If the goal of a definition of a term-forming operator is to assign to the paradigm identity-sentences contents that are evidently logically equivalent with the contents ordinarily expressed by those sentences, then there are multiple, extensionally-inequivalent ways of meeting the goal.

4.5.2 Numbers

Frege's focus on identity-sentences in his treatment of directions is not an idiosyncratic feature of that particular analysis. His view of mathematical objects in general is that they are to be understood via an understanding of their canonical identity-sentences, and that discourse in general about such objects is to be analyzed in terms of an analysis of those identity-sentences. Summing up the procedure he's followed in *Grundlagen* with respect to cardinal numbers, Frege says:

> Now for every object there is one type of proposition which must have a sense, namely the recognition-statement, which in the case of numbers is called an identity. Statements of number too are, we saw, to be considered as identities. The problem, therefore, was now this: to fix the sense of a numerical identity, that is, to express that sense without making use of number words or the word "number." The content of a recognition-judgment concerning numbers was found to be this, that it is possible to correlate one to one the objects falling under a concept *F* with those falling under a concept *G*. Accordingly, our definition had to lay it down that a statement of this possibility means the same as a numerical identity.[36]

If we get the identity-sentences right, Frege seems to think, then we have the materials for giving an adequate analysis of discourse about numbers. Similarly for other mathematical objects. At *Grundlagen* §60, Frege discusses infinitesimals as follows:

> It is enough if the proposition taken as a whole has a sense; it is this that confers on its parts also their content.
>
> This observation is destined, I believe, to throw light on quite a number of difficult concepts, among them that of the infinitesimal,* and its scope is not restricted to mathematics either. [footnote:] The problem here is not, as might be thought, to produce a segment bounded by two distinct points whose length is *dx* but rather to define the sense of an identity of the type *df(x) = g(x)dx*.[37]

And at §104:

> In the same way with the definitions of fractions, complex numbers and the rest, everything will in the end come down to the search for a judgment-content which can be transformed into an identity whose sides precisely are the new numbers. In other words, what we must do is fix the sense of a recognition-judgment for the case of these numbers.[38]

With respect to the central case in *Grundlagen*, the case of cardinal numbers, Frege's view is that in order to "define the sense of a proposition in which a number word occurs,"[39] we must do as follows:

> In our present case, we have to define the sense of the proposition
>
> 'the number which belongs to the concept F is the same as that which belongs to the concept G';
>
> that is to say, we must reproduce the content of this proposition in other terms, avoiding the use of the expression
>
> 'the number which belongs to the concept F'.
>
> In doing this, we shall be giving a general criterion for the identity of numbers.[40]

The relevant recognition-judgments, i.e., those of the form:

> "The number which belongs to the concept *F* = the number which belongs to the concept *G*,"

or, as we'll abbreviate it,

> (N) *NxFx = NxGx*

are to be analyzed in just the way that identity-talk concerning directions was analyzed, which is to say that the analysis is to provide for each instance of (N)

an analysans-sentence that (1) is an identity, and (2) preserves the equivalence between that instance of (N) and the corresponding instance of the statement-type that, as Frege understands it, explains the content of a numerical identity. This latter is, as Frege puts it,

> "There exists a relation ϕ which correlates one to one the objects falling under the concept *F* with the objects falling under the concept *G*,"[41]

or, as we'll abbreviate it,

(EQ) $Fx \approx Gx$.

Frege's way of meeting this requirement is simply to say that "the number which belongs to the concept *F*" is to be analyzed as "the extension of the concept *equinumerous with the concept F*," where "equinumerous with" is the relation noted above that holds between *F* and *G* iff there is a one-one correlation between the objects falling under *F* and those falling under *G*.

The satisfaction of conditions (1) and (2) is straightforward: (N), under the proposal, is analyzed as

(N_{Gl}) $\quad \text{ext}(\xi \approx Fx) = \text{ext}(\zeta \approx Gx)$.

Given the Fregean assumption of the (second-level) principle of extensionality,[42] (N_{Gl}) is easily inter-derivable with (EQ). That is to say, if (N) is analyzed as (N_{Gl}), then the analysis meets the Fregean criteria. The demonstration of this is the content of *Grundlagen* §73, in which Frege sketches the proof of (N_{Gl}) from (EQ) and notes that the proof of the converse is straightforward.

The same can be said for the *Grundgesetze* analysis of (N), under which numbers are understood not as the extensions of second-level equinumerosity-concepts but as the extensions of closely-related first-level concepts. In particular, the number that belongs to the concept *F* is, according to *Grundgesetze*, understood as the extension of that concept under which fall all and only the extensions of concepts equinumerous with *F*:

(N_{Gg}) $\quad \grave{\epsilon}(\exists H(\epsilon = \grave{\alpha}H(\alpha) \,\&\, H \approx F)) = \grave{\epsilon}(\exists H(\epsilon = \grave{\alpha}H(\alpha) \,\&\, H \approx G))$

Appealing in this case just to the first-level extensionality principle (now Axiom V), it's straightforward to see that (N_{Gg}) is also inter-derivable with (EQ).[43] If the goal is to provide an analysis of (N) that preserves its identity-statement syntax and that is demonstrably logically equivalent with (N)'s ordinary content as given by (EQ), it is clear that there are infinitely many extensionally-inequivalent ways to do this. Preservation of the reference of each term of the

form "the number of *F*'s" is no part of what's required in order to give an adequate analysis, along Fregean lines, of statements of the form "the number of *F*'s = the number of *G*'s."

It is not difficult to see the appeal of this strategy. Focusing just on statements of the form (N), the Fregean picture is that a derivation of an instance of (N_{Gl}) or of (N_{Gg}) from premises *P* would show that the ordinary content of (N) itself follows logically from that of *P*. This "transfer" of logical grounding doesn't require that the sentences (N), (N_{Gl}), and (N_{Gg}) express the same content but merely that the inference from (N_{Gl}) to (N) requires just (1) the easy logical inference from (N_{Gl}) to (EQ), relying on (PE); and (2) the analytic claim that (EQ) offers a reasonable analysis of the content of (N), as ordinarily understood. Similarly for the inference from (N_{Gg}) to (N). If one wanted to establish that a particular instance of (N) was grounded in logic, it would suffice to prove the corresponding instance of (N_{Gl}) or of (N_{Gg}) from principles of pure logic. That is to say, it would have sufficed if Frege had been right about (1) and (2).

4.5.3 Arithmetic

Frege's analyses of 0 and of successor flow naturally from his account of (N). In keeping with the idea that cardinal numbers are to be understood as essentially the cardinalities *of* first-level concepts, Frege explains 0 as the cardinal number that belongs to the concept *not self-identical*, and explains successor as that relation that holds between *n* and *m* if for some *F* and *x*: *x* falls under *F*, *n* is the number that belongs to *F*, and *m* is the number that belongs to the concept ($Fy \,\&\, y \neq x$). The cardinal number 1, i.e. the successor of 0, is then the number that belongs to the concept ($y = 0$); 2 is the number that belongs to the concept ($y = 0$ or $y = 1$); and so on.

Despite their initial unfamiliarity, the close connection between these accounts of the numbers and an ordinary understanding of applied numerical discourse is illustrated in *Grundlagen* by the following considerations. First of all, for any concept *F*, Frege's rendering of

(App 0) 0 = the number that belongs to the concept *F*

is provably equivalent,[44] using just Fregean laws of logic, with the ordinary

(App* 0) $\forall x \sim Fx$

Frege's demonstration of this fact is the content of *Grundlagen* §75, titled "Zero is the number which belongs to a concept under which nothing falls. No object falls under a concept if zero is the number belonging to that concept."[45]

Similarly, it is a straightforward matter to prove the content of

(App 1) 1 = the number that belongs to the concept *F*,

as Frege analyzes it, from the ordinary

(App* 1) $\exists x(Fx \,\&\, \forall y(Fy \rightarrow y = x))$,

and vice-versa. The step-by-step demonstration of the equivalence between (App 1) as analyzed in *Grundlagen* and (App* 1) is the content of *Grundlagen* §78, propositions 2–4.

Frege provides parallel demonstrations in *Grundgesetze*: the *Grundgesetze* version of (App 0) is demonstrated at Vol 1 §§96–99 to be inter-derivable with (App* 0); the inter-derivability of *Grundgesetze*'s version of (App 1) with (App* 1) is shown at §§102–103 and 106–107.

Frege's rendering in each case of

(App succ) The successor of *n* = the number that belongs to the concept *F*

is, similarly, provably equivalent with

(App* succ) $\exists x(Fx \,\&\, n$ is the number that belongs to $(Fy \,\&\, y \neq x))$,

i.e. with, as one might put it, the claim that there's "one more" than *n* things falling under *F*.

Similarly in the other direction. These proof-theoretic equivalences hold whether "the number that belongs to the concept *F*" is analyzed as in *Grundlagen* or as in *Grundgesetze*. As a result, for any finite *n*, Frege's version of "the successor of . . . the successor of 0 belongs to the concept *F*," where "successor of" is iterated *n* times, will be straightforwardly inter-derivable with the usual first-order rendering of "there are exactly *n F*'s."

Frege takes the finite cardinals themselves to be those objects that bear the weak ancestral (as defined in *Begriffsschrift*) of the successor relation to 0.[46] As he demonstrates sketchily in *Grundlagen* §§78-83, the fundamental arithmetical truths come out right on the proposed analysis, which is to say that they are provable from purely logical laws once 0, successor, and finite cardinal are understood in *Grundlagen's* way. In these sections, Frege outlines demonstrations that the successor relation on the finite cardinals is 1-1, that its strong ancestral is irreflexive and transitive on this collection, that each finite cardinal other than 0 is the successor of some finite cardinal, and that each finite cardinal has a successor.[47] As he notes in §80, the principle of induction on the finite cardinals falls out immediately from the definitions. That is to say, what we get here is essentially a sketch of how to demonstrate the Dedekind-Peano axioms. (Parallel demonstrations with the details filled in are provided in *Grundgesetze* Vol 1.)

Having reached this point in *Grundlagen*, Frege takes himself to have said (and shown) enough to make it plausible that his account of the finite cardinals is a good one. We can follow his reasoning to this point by going over his own summary of the *Grundlagen* analysis as given in §§106–109. The summary opens as follows, covering some ground already quoted above:

> §106 Let us cast a final brief glance back over the course of our enquiry. After establishing that number is neither a collection of things nor a property of such, yet at the same time is not a subjective product of mental processes either, we concluded that a statement of number asserts something objective of a concept. We attempted next to define the individual numbers 0, 1, etc., and the step from one number to the next in the number series. Our first attempt broke down, because we had defined only the predicate which we said was asserted of the concept, but had not given separate definitions of 0 or 1, which are only elements in such predicates. This resulted in our being unable to prove the identity of numbers. It became clear that the number studied by arithmetic must be conceived not as a dependent attribute, but substantivally. Number thus emerged as an object that can be recognized again, although not as a physical or even a merely spatial object, nor yet as one of which we can form a picture by means of our imagination. We next laid down the fundamental principle that we must never try to define the meaning of a word in isolation, but only as it is used in the context of a proposition; only by adhering to this can we, as I believe, avoid a physical view of number without slipping into a psychological view of it. Now for every object there is one type of proposition which must have a sense, namely the recognition-statement, which in the case of numbers is called an identity. Statements of number too are, we saw, to be considered as identities. The problem, therefore, was now this: to fix the sense of a numerical identity, that is, to express that sense without making use of number words or the word 'number.' The content of a recognition-judgment concerning numbers we found to be this, that it is possible to correlate one to one the objects falling under a concept *F* with those falling under a concept *G*. Accordingly, our definition had to lay it down that a statement of this possibility means the same as a numerical identity.[48]

That is to say: the account is to rest on its treatment of statements of the form (N), and a condition of adequacy is that (N) be analyzed in a way that preserves its equivalence with (EQ). Frege next summarizes the argument to the effect that the analysis of (N) must also deliver identity-sentences, and rehearses his strategy for meeting both of these requirements:

> One doubt, however, still remained, which was this. A recognition-statement must always have a sense. But now if we treat the possibility of correlating one to one the objects falling under the concept *F* with the objects falling under the concept *G* as an identity, by putting for it "the Number which belongs to the concept *F* is identical with the Number which belongs to the concept *G*," thus introducing the expression "the number which belongs to the concept *F*," this gives us a sense for the identity only if both sides of it are of the form just mentioned. A definition like this is not enough to enable us to decide whether an identity is true or false if only one side of it is of this form. We were thus led to give the definition:
>
> > The Number which belongs to the concept *F* is the extension of the concept "concept equinumerous with the concept *F*,"
>
> where a concept *F* is called equinumerous with a concept *G* if there exists the possibility of one-one correlation referred to above.
>
> In this definition the sense of the expression 'extension of a concept' is assumed to be known. This way of getting over the difficulty cannot be expected to meet with universal approval, and many will prefer other methods of removing the doubt in question. I attach no decisive importance even to bringing in the extensions of concepts at all.[49]

The "difficulty," again, is that if we analyze (N) directly as (EQ), we will fail to have given sense to ordinary arithmetical identity-sentences. We need instead to analyze (N) via an identity-sentence, one with singular terms on each side of the identity-sign, since only by doing so will we be treating numbers as objects. And only by so treating them can we make sense, from Frege's point of view, of arithmetical discourse in general, including quantification over numbers. Hence the analysis of (N) as (N_{Gl}), since this is one way to meet both criteria: to analyze (N) (i) in a way that reflects the closeness of (N)'s content with that of (EQ), and (ii) as an identity. Similarly for the analysis of (N) as (N_{Gg}).

We can now see that Frege's remark that he "attach[es] no decisive importance even to bringing in the extensions of concepts at all," and the fact that he claims no unique correctness for his account, is entirely to be expected, given his analytic goals. There's no reason to read the comment as indicating that he is uninterested in providing analysans-sentences whose contents are close to the ordinary contents of arithmetical sentences. Quite the contrary: the remark comes, as we see, in the course of a rehearsal of detailed philosophical argument regarding the ordinary meanings of arithmetical terms—e.g. that they are not about ideas or about physical objects, that the truths they express are not subjective, and so on. For Frege, what's important to "get right" in the analysis is the account of the contents of whole sentences, especially of identity-sentences, and not the referents of singular terms.

Frege's summary of his reduction continues:

> It now still remained to define one-one correlation; this we reduced to purely logical relationships. Next, we first gave an outline of the proof of the proposition: the number which belongs to the concept *F* is identical with the number which belongs to the concept *G*, if the concept *F* is equinumerous with the concept *G*; and then gave definitions of zero, of the expression "*n* follows in the natural series of numbers directly after *m*," and of the number 1, showing that 1 follows in the natural series of numbers directly after 0. After adducing a number of propositions which can easily be proved at this stage, we proceeded to go rather more closely into the following proposition, from which we learn that the number series is infinite:
>
> After every number there follows in the natural series of numbers a number.
>
> . . . [W]e succeeded in showing that the inference from *n* to (*n* +1), which is ordinarily held to be peculiar to mathematics, is really based on the universal principles of inference in logic.[50]

That is to say: Frege sketches the proof of the equivalence of (N_{Gl}) with (EQ), defines 0 and successor, sketches the demonstration that 1 is the successor of 0, and "adduc[es] a number of propositions which can easily be proved at this stage." These include the equivalence of (App 1) and (App* 1) and essentially the Peano axioms, with sketches given of the demonstration of the infinity of the finite cardinals and of the principle of induction. He then concludes that

> From all the preceding it thus emerged as a very probable conclusion that the truths of arithmetic are analytic and a priori; and we achieved an improvement on the view of Kant.[51]

In short, the *Grundlagen* reduction involves:

1. An extended argument to the effect that the crucial contents that need to be analyzed in order to give an account of the nature of cardinal numbers are those propositions expressed by instances of (N);
2. An argument that each instance of (N) should be understood to express essentially what its corresponding instance of (EQ) expresses;
3. An argument that the right way to analyze instances of (N) is to respect #2 above while providing analysans-sentences that take the form of identities;
4. The proposal to analyze each instance of (N) via the corresponding instance of (N_{Gl}), noting that this proposal meets the conditions of adequacy just noted, but without any claim to uniqueness;

5. The proposal to understand 0 and 1 as the numbers belonging respectively to the concepts *not self-identical* and *identical with 0*. Along with this proposal comes the demonstration that, with 0 and 1 thus understood, Frege's analyses of (App 0) and (App 1) deliver contents that are demonstrably logically equivalent with (App* 0) and (App* 1) respectively, i.e. with straightforwardly good analyses of the ordinary contents of (App 0) and of (App 1);
6. The proposals to understand successor as given in (App* Succ), and to understand the finite cardinals as those objects bearing the weak ancestral of the successor operation to 0;
7. The (sketches of) demonstrations that, so understood, 0, 1, successor and the concept *finite cardinal* have all of the usual arithmetical properties required in order to characterize the series of finite cardinals in their usual ordering.

And this, as Frege knows, is enough for arithmetic. It's a straightforward matter, from here, to define the usual arithmetical operations in terms of 0 and successor, so that, with respect to the arithmetic of the finite cardinals, there is nothing else to "get right."

Let's return to the problematic example of (β), "1 is the successor of 0." As above, the *Grundlagen* analysis of (β) is roughly:

> (β*) For some concept *F* and some object a falling under *F*, the extension of the concept *equinumerous with the concept identical with the extension of the concept (equinumerous with the concept (not self-identical))* is the extension of the concept *equinumerous with the concept F*, and the extension of the concept *equinumerous with the concept not self-identical* is the extension of the concept *equinumerous with the concept falling under F, but not identical with a*.

To repeat our earlier concern, it's not plausible to claim that Frege's analysis of (β) is sense-preserving, at least if senses are individuated along the lines of the cognitive criterion. And Frege clearly did not intend his analysis to preserve the ordinary reference of the two numerals that appear in (β). We want to know in what sense it can be claimed, under these conditions, that Frege intended to offer an analysis of, rather than a model or replacement of, the content of this and other sentences of ordinary arithmetical discourse.

The answer suggested by Frege's procedure is that the adequacy of an analysis of (β) turns on the adequacy of its treatment of the fundamental arithmetical notions in terms of which (β), as ordinarily understood, is straightforwardly explained. The important point is that (β*) is provably logically equivalent with a good analysis of the ordinary content, as Frege understands it, of (β). There are two ways to see that Frege's reduction, had it not had internal difficulties, would have satisfied this condition. The first is to notice that an analysis of (β) is

adequate if it's delivered by an adequate analysis of 0, of 1, and of successor. And the adequacy of the Fregean analysis of these fundamental notions is given by the satisfaction of the conditions noted above. The second way to see the adequacy of (β*) is to note that for Frege: 0 is essentially the number belonging to any concept under which nothing falls (as expressed in (App* 0)); 1 is essentially the number belonging to any concept under which exactly one thing falls (as expressed in (App* 1)), and successor is that relation that holds between two numbers if the first is the number belonging to a concept under which falls one more object than fall under a (or any) concept to which the second number belongs. Hence an analysis of (β) is adequate if and only if the analysans is straightforwardly logically equivalent with a claim to the effect that

$$\forall F \forall G[((\exists x)(Fx \,\&\, \forall y(Fy \rightarrow y = x)) \,\&\, \forall z{\sim}Gz) \rightarrow \exists H \exists w(Hw \,\&\, NxFx = NxHx \,\&\, NxGx = Nx(Hx \,\&\, x \neq w))].$$

That is, as one might more colloquially put it, if there's one F and 0 G's, then the number of F's bears to the number of G's that relation expressed in (App* succ). And (β*), despite its long-windedness, clearly meets this condition.

Though Frege does not define addition in *Grundlagen*, a canonical recursive definition of the operation will have the result that contents like that of "3 + 2= 5" are understood in terms of the fundamental notions already analyzed, and hence that the adequacy of the treatment of these concepts turns on the adequacy of what *Grundlagen* has already given us. Similarly for the rest of the arithmetic of the finite cardinals.

The adequacy of the Fregean analysis, again leaving aside its reliance on the false assumptions regarding extensions or courses-of-values, would not have sufficed for the truth of his logicist claim. All that follows from this adequacy is the all-important but still preliminary claim that *if* Frege's fully-analyzed versions of arithmetical truths had been provable from purely logical principles, then so too would be the truths of arithmetic themselves, i.e. those thoughts expressed in ordinary arithmetical discourse.

4.5.4 Alternative Reductions

Despite the difficulties with Frege's analyses, to be discussed briefly below, one thing that's clear is that, given Frege's understanding of conceptual analysis, the multiple reducibility of the numbers to extensions or to courses of value is entirely consistent, from Frege's point of view, with the idea that each of these multiple reductions can in principle provide a faithful analysis of arithmetical discourse. For given Frege's understanding of what's crucial in order successfully to "reproduce the content" of ordinary arithmetical statements, such content-reproduction clearly doesn't require term-by-term reference preservation.[52]

That there is room for arbitrariness in Frege's reference-fixing procedure is due to the fact that there are multiple equally-good functions *f* from concepts to objects such that for every *F* and *G*, the identity

(ID_f) $\quad f(F) = f(G)$

is easily and obviously logically equivalent with its counterpart equinumerosity claim (EQ). Any such function *f* will do for Frege's purposes, which is to say that for any such *f*, its range forms a collection well-suited to serve as the referents of numerical terms. The members of such a collection are well-suited in the sense that identity-statements involving numerals so understood express contents or thoughts logically equivalent with the equinumerosity-claims that Frege takes numerical identity-statements, as ordinarily understood, to express.

But it's equally clear that not just any appropriately-structured collection of objects would do for Frege's purposes. Consider for simplicity again just the case of the finite cardinals. Supplied with any ω-sequence $\langle a_0, a_1, \ldots \rangle$, one can correlate concepts with members of the sequence in such a way that all and only the concepts under which nothing falls are correlated with a_0, and a concept *F* is correlated with a_{n+1} iff there's an object *b* falling under *F* and such that the concept "falling under *F* but not identical with *b*" is correlated with a_n. It will then be the case that for each concept *F* and natural number *n*, *F* is correlated with a_n iff there are exactly *n F*'s. But this is as close as such an arbitrary reduction will get to a Fregean reduction. The claim that a_n is correlated with *F* will in general be only materially, and not logically, equivalent with the ordinary claim that there are *n F*'s. In particular, if the existence of the sequence $\langle a_0, a_1, \ldots \rangle$ is grounded in principles of e.g. spatial intuition or set theory, then no appeal to laws of logic will enable one to prove the "reduced" from the ordinary claim.[53] Similarly, the "reduced" version of (N) will not be logically equivalent with (EQ). If Frege had been right, however, and if his logic had not succumbed to Russell's paradox, the situation would have been quite different with his own reduction. For Frege was in a position to argue that the claims of ordinary arithmetic can be properly analyzed in such a way that they are logically equivalent with the *Grundlagen* versions of those claims, and hence that the grounds of the latter suffice to ground the former.

4.6 Conclusions

The crucial difficulty in Frege's reduction is simply the failure of his Principle of Extensionality and, together with it, the axiomatic foundation of *Grundgesetze*. By revealing that there is no function *ext* from concepts to objects such that contents of the form

$$\boldsymbol{ext(F) = ext(G)}$$

are logically equivalent with their counterparts of the form

$$\boldsymbol{\forall x(Fx \text{ iff } Gx)},$$

Russell's paradox shows that Frege's whole analytic procedure is a disaster, even prior to the difficulties with his formal proofs. For there's now no way to satisfy the demand that each statement of the form (N) be analyzed via a statement whose content is (1) an identity, and (2) logically equivalent with the corresponding instance of (EQ). Similarly, no identity-statement can meet the condition for analyzing (App 0) that its content be logically equivalent with that of (App* 0). And so on. For the existence of identity-statements logically equivalent with the corresponding universal generalizations requires objects whose existence is guaranteed by principles of pure logic—i.e. of, as Frege calls them, "logical objects." And the lesson of Russell's Paradox, as Frege sees it, is that there are no such things.[54] The only options left for an analysis of arithmetical discourse are analyses that reject the idea that numbers are objects and those, toward which Frege leans at the end of his life, that preserve the account of arithmetic as a science of objects but that turn to a non-logical source—in this case, geometry—for the nature and the existence of those objects.

For our purposes, i.e. for the purpose of investigating Frege's conception of logic, the important conclusion to draw from his failed logicist project is that Frege should be taken at his word when he claims to be investigating the logical grounds of the truths of arithmetic, those truths that are under investigation during the course of ordinary arithmetical research and are expressed by the sentences of ordinary arithmetical discourse. As Frege characterizes the project in 1897, in a comparison between his methodology and Peano's, we see once again the role of analysis in answering Frege's foundational questions:

> I became aware of the need for a conceptual notation when I was looking for the fundamental principles or axioms upon which the whole of mathematics rests. Only after this question is answered can it be hoped to trace successfully the springs of knowledge upon which this science thrives. Even if this question belongs largely to philosophy, it must still be regarded as mathematical. The question is an old one: apparently it was already being asked by Euclid. If it has nevertheless not yet been answered satisfactorily, the reason is to be sought in the logical imperfection of our languages. . . . For an investigation such as I have in mind here it is not sufficient for us just to convince ourselves of the truth of a conclusion, as we are usually content to do in mathematics; on the contrary, we must also be made aware of what it is that justifies our

> conviction, and upon what primitive laws it is based. For this are required fixed guiding-lines, along which the deductions are to run; and in verbal languages these are not provided. . . .
>
> . . .
>
> What I am aiming for, then, is uninterrupted rigor of demonstration and maximal logical precision, together with perspicuity and brevity.
>
> I cannot so definitely specify what aim it is that Peano is pursuing with his conceptual notation or mathematical logic: I have to rely largely upon conjecture. . . . This much I think I can gather . . . that an examination of the foundation of mathematics is not what initiated it—nor has it been a determinant for its mode of execution. For straightaway, in §2 of his *Introduction*, brief tags are introduced for the classes of the real numbers, the rational numbers, the prime numbers, etc., which means that all these concepts are assumed as already familiar. The same thing happens with the meanings of the operation-signs '+', '-', '×', '√', etc., from which it is to be gathered that an analysis of these logical structures into their simple components was not the intention. And since, without such an analysis, an investigation like the one I projected is impossible, such an investigation could not have been among Mr. Peano's intentions.[55]

Henceforth, we take Frege to mean what he says when he says that the proof of his highly-analyzed versions of arithmetical truths will suffice to establish the logical grounds of the arithmetical truths themselves.

Frege's failure to provide a readily-verifiable criterion for the adequacy of analysis means that part of his procedure of investigating logical relations, specifically the "analytic" part, will, in each instance, be open to philosophical debate. Claims to the effect that proposed analyses are adequate, e.g., claims about the accuracy of understanding (N) in terms of (EQ), will be claims about which there can in principle be disagreement, and disagreement for whose solution there is no algorithm. But this failure does not mean that the adequacy of particular analyses is either unknowable or insusceptible to philosophical argument. And that a particular analysis of a given thought is in fact adequate is what we require in order to apply Frege's procedure to the investigation of the logical status of that thought. Frege's lack of clarity about the criteria for adequate analysis would seem to indicate simply that he, like us, takes analysis to be a philosophically demanding procedure, one for which it is not a straightforward matter to give a recipe, but which is sometimes clearly successful.

The role of conceptual analysis in Frege's logicist project can be summed up as follows. A thought τ, expressed by a relatively-simple sentence s, is subjected to an analysis which terminates in the thought τ^*, itself expressible by a

relatively-complex sentence s^*. The latter sentence is derivable from a set Σ of sentences from which the original s is not so derivable. The derivation shows (or would show, if Frege's logic weren't problematic) that τ^* is logically entailed by the thoughts expressed by the members of Σ; the connection between τ^* and τ (namely, that τ^* is the result of a successful analysis of τ) shows that τ itself is similarly logically entailed by those thoughts.

The idea that conceptual analysis can be informative about logical relations is to some extent a familiar one, one that plays a significant role in much of what we do both within and outside of mathematics. The familiar analyses of such mathematical notions as *continuous function*, *infinite set*, *circle*, etc. are crucial to the proofs we give of every nontrivial claim about them. Similarly, though more contentiously, for such non-mathematical notions as those of *knowledge, free will*, and so on. But the implications of this core idea, when combined with the use of formal systems to demonstrate relations of logical entailment, are less familiar. We turn in the next three chapters to the implications of the Fregean approach for the role of formal systems in the demonstration of such logical relations as entailment and consistency, and for the assessment of those systems themselves.

Notes

1. [1914] p 228/211.
2. *Grundlagen* (hereafter *Gl*) §1. See Paul Benacerraf [1981] for a discussion of the extent to which Frege's analytic project follows in the footsteps of these earlier analyses. But see below (§4.3) for disagreement with Benacerraf's understanding of what that project involves.
3. *Gl* §2.
4. See, e.g., Frege's [1881] characterization of the content of *Begriffsschrift* formulas via natural-language expressions, e.g., "12 is a multiple of 4," "13 is prime," etc.; here the contents of the numerals are left unanalyzed, while that of "is a multiple of" and "prime" are analyzed in standard Fregean ways. See Frege's characterization of *Grundgesetze*'s goal at *Grundgesetze* (hereafter *Gg*) 1 §0 as that of the "derivation of the simplest laws of Numbers by logical means alone."
5. [1914] p 226/209. In the next paragraph, Frege seems to count the difficulty of judging whether an analysis is correct as just the difficulty of judging whether the thought is preserved.
6. [1914] p 228/211.
7. Ibid.
8. *Gl* Introduction p vii.
9. This is to oversimplify slightly. A speaker who fails to understand a sentence does not necessarily fail to grasp the sense it expresses; she can simply fail to know that that sense is expressed by the given sentence. This is the case with failing to understand the sentences of a foreign language. In the case of simple arithmetical sentences, however, one can't plausibly maintain that ordinary speakers grasp the senses of those sentences while failing to associate the senses with the sentences. For outside of the context of learning arithmetic, via the use of those very sentences, such a speaker has no access to those

senses. If, in the ordinary run-of-the-mill cases, speakers fail to understand the rudimentary arithmetical sentences of their home language, they will in such cases fail to grasp the senses expressed by those sentences.

10. This is not to say that one can never apply such a criterion. Given, e.g., two sentences of a reliable formal system, one might well be in a position to know that they express logically-equivalent thoughts in virtue of the interderivability of the sentences without any antecedent knowledge of the adequacy of one as an analysis of the other. But this is not the case with the ordinary analysandum-sentences and their formal Fregean analysantia.
11. Frege himself seems to suggest such a criterion. E.g. in a response to a complaint of Husserl's that his analyses do not preserve something like sense, Frege responds that "what matters to [mathematicians] is the thing itself: the reference of the words" [1894] pp 319–320/200. For discussion of this passage, see Dummett [1991c].
12. See, e.g., Benacerraf [1981]; Weiner [1984], [2007]; Dummett [1991a] Ch. 14.
13. *Gl* §107.
14. As he says at *Gg* 1 Introduction p x/6: "we just cannot get on without" courses-of-values (extensions).
15. See Benacerraf [1981].
16. See Weiner [1984].
17. Weiner [2007] p 690. See also the criterion as expressed in Weiner [1990] pp 123–124:

> . . . the constraint on the replacement of pre-Fregean sentences with Fregean sentences is that the pre-Fregean inferences that were accepted by mathematicians must be constructible in Fregean arithmetic.

18. The reason for saying "apparent" is that Weiner takes Frege to deny that arithmetical sentences as ordinarily understood express truths or falsehoods.
19. *Gl* §7.
20. As Frege puts it at *Gl* §14:

> For purposes of conceptual thought we can always assume the contrary of some one or other of the geometrical axioms, without involving ourselves in any self-contradictions when we proceed to our deductions, despite the conflict between our assumptions and our intuition. The fact that this is possible shows that the axioms of geometry are independent of one another and of the primitive laws of logic, and consequently are synthetic. Can the same be said of the fundamental propositions of the science of number? Here, we have only to try denying one of them, and complete confusion ensues. Even to think at all seems no longer possible. The basis of arithmetic lies deeper, it seems, than that of any of the empirical sciences, and even than that of geometry.

21. *Gl* §9.
22. *Gl* §12.
23. *Gl* §19.
24. *Gg* 1 Introduction p viii/5; emphasis added.
25. Dummett [1991a] p 179.
26. *Gl* §91. See also *Begriffsschrift* (hereafter *Bff*) Preface p iii/5; [1897b] p 362/235; *Gg* 1 Introduction p vii/3.
27. Heck [1993].
28. The account of the finite cardinals in *Grundlagen* is wrapped up with Frege's accounts of cardinal number in general and of the first infinite cardinal. We look here just at the finite cardinals for the sake of simplicity.
29. For an enlightening discussion of the mathematical setting of Frege's treatment of this issue, see Wilson [1992].
30. *Gl* §64.
31. *Gl* §64.

32. *Gl* §68.
33. [1914] p 227/210.
34. Frege doesn't state (PE) in *Grundlagen*, but simply presumes it in all of his proof-sketches involving extensions. (PE) is the heart of what is later codified as Axiom V of *Grundgesetze*.
35. This requires again the assumption just noted that *parallel* is an equivalence relation. Things are considerably smoother when we turn to the definition of number, where the fact that equinumerosity is an equivalence relation is proven rather than assumed.
36. *Gl* §106.
37. *Gl* §60.
38. *Gl* §104.
39. *Gl* §62.
40. Ibid.
41. *Gl* §72.
42. i.e. that for all second-level concepts Π and Ψ: ext($\Pi(\xi)$) = ext($\Psi(\zeta)$) iff $\forall$H(Π(H) iff Ψ(H)).
43. See *Gg* 1 §§54–64.
44. For (App 0) is shorthand for:

    (1) the extension of the concept *Equinumerous with the concept* <u>*not self-identical*</u> = the extension of the concept *Equinumerous with the concept* F;

    Appeal to (PE) and the definition of equinumerosity yield:

    (2) there's a 1–1 map from the objects falling under *F* to the non-self-identical objects; from which (App* 0) follows easily. Similarly in reverse.
45. *Gl* §75 and Table of Contents p xi.
46. An object *a* bears the weak ancestral of R to *b* iff either *a* = *b*, or *a* bears the ancestral (as defined in *Begriffsschrift*) of R to b.
47. See Boolos and Heck [1998] for a discussion of difficulties with the last proof-sketch.
48. Gl §106.
49. Gl §107.
50. Gl §108.
51. Gl §109.
52. Nevertheless, we might ask about the truth-value of such sentences as "0_{Gg} = 0", where the first term is the *Grundgesetze* zero, and the second is the zero of ordinary discourse. The answer is that this sentence has no truth-value. The ordinary "0" is not a term of L_G, and the semantic stipulations Frege gives for L_G determine the reference of 0_{Gg} only to the extent necessary for linguistic completeness. If we were to add the ordinary term to the language L_G, then, as Frege says in *Gg* 1 §10, we would have to provide further stipulations to handle the newly-well-formed sentences of the language. With respect to *Grundlagen*, the situation is less clear, because Frege is less clear about the status of reference to extensions in 1884 than one might like. The most straightforward thing to say is that the answer is the same with respect to *Grundlagen* as it is with respect to *Grundgesetze*: there is no fact of the matter about whether the ordinary numerals and the *Grundlagen* numerals co-refer. This view about Fregean object-reference is inconsistent with the view that all functions are everywhere defined. But as we saw in Chapter 3, the grounds for attributing the latter view to Frege are problematic. The clear indeterminacy of the identity-question with respect to *Grundgesetze*'s numerals is further reason to doubt the condition of total definition. Finally: it is important to note that the indeterminacy here applies not just to sentences like "0_{Gg} = 0," but also to sentences like "0_{Gg} = Mt. Aetna." The semantic stipulations given in *Grundgesetze* leave all such identity-sentences undetermined, and hence the strangeness of Frege's position with respect to the *Grundgesetze* numerals cannot be avoided by holding that the ordinary "0" fails to refer.
53. Unless, contra the pre-1920s Frege, the ordinary claim is at bottom geometrical.
54. Regarding the "logical source of knowledge," Frege writes in [1924/25] (pp. 298–299/278–279) that "it seems that this on its own cannot yield us any objects."
55. [1897b] pp 362–365/235–237.

5

Analysis and Consistency: The Case of Geometry

5.1 Introduction

As we have seen in the case of arithmetic, the most familiar ways of expressing a body of truths may not be the most revealing with respect to logical structure. It is only, for example, once we see that the ancestral of a relation can be cashed out in Frege's way that we see how to understand various ordinary claims about natural numbers in terms of those properties had by zero and passed on by the successor relation. It is only once we have understood these claims in such a way, in turn, that we are in a position to see that, on occasion, they are entailed by surprisingly modest premises or that they themselves entail surprisingly rich conclusions.

That relations of logical entailment can in this way be brought to light by conceptual analysis is due simply to the fact, as Frege sees it, that these logical relations are sensitive to the meanings of such terms as "successor," "natural number," and so on. That "every natural number has a successor" expresses a logical truth in Frege's eyes, while e.g. "every quadruped has a neck" does not, is by no means anomalous from Frege's point of view: from that point of view, the fact that two sentences share superficial grammatical structure is no reason to suppose that the thoughts they express share logical properties.

The fact that logical entailment, from Frege's point of view, has to do not just with surface-level grammatical structure but also with the meanings of individual terms means that there is a gap between derivability and entailment: while the derivability in a good formal system of a sentence ϕ from a set Σ of sentences shows that the thought $\tau(\phi)$ expressed by ϕ is logically implied by the set $\tau(\Sigma)$ of thoughts expressed by the members of Σ, the converse is false: the non-derivability, in such a system, of ϕ from Σ is no guarantee that $\tau(\phi)$ fails to follow logically from $\tau(\Sigma)$. As we see in the case of arithmetic, the non-derivability of ϕ from Σ is compatible with the existence of a more-complex sentence ϕ' and set Σ' expressing analyses respectively of what's expressed by ϕ and by Σ, and

such that ϕ' is derivable from Σ'. And in such a case, Frege takes it that the derivation shows $\tau(\phi)$ to follow logically from $\tau(\Sigma)$. In short: derivability in a good formal system demonstrates logical entailment, but the converse is false: non-derivability is no guarantee of a failure of logical entailment.

The gap as Frege understands it between derivability and logical entailment plays an important role in Frege's famous disagreement with Hilbert over consistency proofs in geometry. As we'll see below, Frege's understanding of the logical relations, and his differences with Hilbert, mark him off quite dramatically not just from Hilbert himself but also from the now-standard model-theoretic approach to logical properties and relations that grows out of Hilbert's and related work.

The purpose of this chapter is to take a careful look at the Frege-Hilbert controversy, with an eye toward clarifying the central differences between Frege's conception of entailment and a standard modern conception that finds an early expression in Hilbert's *Foundations of Geometry*. In the succeeding chapter, we take a look at some of the ways in which these differences persist as differences between the Fregean and some relatively modern approaches to the logical relations.

5.2 Frege-Hilbert

In the summer of 1899, at a celebration marking the unveiling of the new Gauss-Weber monument in Göttingen, David Hilbert delivered a lecture on the foundations of geometry, later published by Tuebner under the title "*Grundlagen der Geometrie*" (*Foundations of Geometry*; hereafter *FG*).[2] Hilbert's purpose in the lecture and monograph is to provide a clear and precise set of axioms for Euclidean geometry, and to prove a number of consistency and independence results regarding those axioms and related theorems. The work is a watershed in the development of logic, presenting as it does a remarkably fruitful and systematic application of the then-emerging technique of "reinterpretation" for the demonstration of consistency and independence results.

Shortly after the lecture, Frege acquired a copy of the monograph, writing to Hilbert in December 1899 that there seemed to be serious flaws in its central technique. Hilbert replied, initiating a correspondence on this topic involving just six letters, and lasting until September 1900. Throughout the correspondence, Frege's reservations grow deeper, while Hilbert remains unmoved by his colleague's complaints. Frege continues his discussion of the central issues and his criticism of Hilbert in two series of essays, both titled "On the Foundations of Geometry," published in 1903 and 1906. Despite the growing importance of Hilbert's and related techniques over the course of Frege's lifetime, Frege seems never to have appreciated their significance. As Frege puts it in a letter to Liebmann in July 1900:

> Clear and inventive as it is in many points, I think that [Hilbert's manuscript] is on the whole a failure and in any case that it can be used only after thorough criticism.[3]

There is a clear sense in which Hilbert is the winner in this debate: the standard methodology in use today for demonstrating consistency and independence is a direct outgrowth of Hilbert's method, while Frege's objections have essentially fallen by the wayside. But this is not to say that Frege's worries about Hilbert's technique have been answered. What we see in this debate is a clash between two quite different ways of understanding the fundamental notions of consistency and independence. Understood in Hilbert's way, consistency and independence are clearly demonstrated by Hilbert's own, and by our contemporary, techniques. Understood in Frege's way, as we'll see, consistency and independence are not demonstrable by these techniques.

5.3 Hilbert's Method

Hilbert's consistency demonstrations in *FG* are all demonstrations of relative consistency, which is to say that in each case the consistency of a set Σ of geometric sentences is reduced to that of a familiar background theory B, demonstrating that Σ is consistent if B is. The important technique Hilbert employs is the *reinterpretation* of the geometric terms appearing in Σ in such a way that, as reinterpreted, the members of Σ express theorems of B. For example, Hilbert's first consistency-proof interprets the terms "point," "line," and "lies on" as standing, respectively, for a particular collection of ordered pairs of real numbers, for a collection of ratios of real numbers, and for an algebraically-defined relation between such pairs and ratios. Under this reinterpretation, the geometric sentences in question express theorems of the background theory of real numbers.

That such a reinterpretation strategy guarantees relative consistency can be seen via the following reasoning: If the set Σ were inconsistent, then it would logically imply a contradiction. Since logical implication is independent of the specific meanings of such terms as "point" and "line," Σ would continue to imply a contradiction under its reinterpretation.[4] But that is just to say that a set of theorems of B would imply a contradiction, hence that B itself would be inconsistent. As Hilbert puts it, "Every contradiction in the consequences of the [geometric sentences] would therefore have to be detectable in the arithmetic of the field Ω."[5]

Independence is demonstrated in exactly the same way. To show that a sentence ϕ is independent of a set Σ of sentences (relative to the consistency of B), one interprets the geometric terms of Σ in such a way that the members of Σ, as

interpreted, express theorems of B, while ϕ expresses the negation of a theorem of B. That is, the independence of ϕ from Σ (relative to the consistency of B) is demonstrated by proving the consistency of $\Sigma \cup \{\sim\phi\}$ relative to that of B.

A crucial feature of Hilbert's approach is its focus not on particular geometrical concepts like *point* and *line* but instead on the logical relations that are said, by the sentences in question, to hold between those concepts. The question of whether the parallels axiom is independent of the other Euclidean axioms has, in Hilbert's hands, entirely to do with the logical structure exhibited by the relevant axiom-sentences, and nothing to do with whether it is geometric *points* and *lines* one is talking about, or some other subject-matter altogether. As Hilbert says:

> [I]t is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep . . . and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things. In other words: any theory can always be applied to infinitely many systems of basic elements.[6]

The consistency- and independence-questions with which Hilbert is concerned are, in short, questions that turn on the general structural features of the sentences in question and not on the particular contents of the geometrical terms appearing in those sentences. On this way of understanding them, consistency and independence have essentially to do with the relation of derivability between sentences, where derivability can be understood in terms of the usual kinds of topic-neutral, syntactically-specifiable rules of inference of a good formal system.[7] A set Σ is consistent if no contradiction is derivable from it via the usual topic-neutral rules; ϕ is independent of Σ if ϕ is not derivable, in this sense, from Σ. As we'll put it in what follows, a Hilbert-style consistency- or independence-proof shows (assuming the consistency of the background theory) that the set Σ is *syntactically consistent*, or that the sentence ϕ is *syntactically independent* of Σ.

One result of this conception of consistency and independence is that, when considering the question of the consistency of a set Σ of sentences, it makes no difference whether one takes Σ's geometric terms (1) to be empty place-holders, (2) to have their ordinary geometric meanings, or (3) to have some particular alternative content, e.g. of the kind assigned by one of Hilbert's interpretations. On Hilbert's understanding of consistency, if Σ is consistent when its terms are interpreted in one way (or simply uninterpreted), then it's consistent under any (other) interpretation. To put this point in Fregean terms: the question of which *thoughts* Σ's members express, or whether they express any thoughts at all, is, with respect to Hilbert's conception of consistency, irrelevant to the question of the consistency of the set.

A second result of Hilbert's approach is that, by treating the geometric terms in Σ as empty place-holders, there is a straightforward sense in which Σ provides a definition. Specifically: where there are *n* geometric terms in Σ, Σ defines an *n*-place relation R_{Σ} satisfied by just those *n*-tuples that, when assigned in order to Σ's terms, make the members of Σ true.[8] The consistency-question treated by Hilbert—namely, that of the set Σ of sentences—is closely related to a consistency-question regarding the relation R_{Σ}, which we can illustrate as follows. Consider for example the two sets of sentences:[9]

(Σ1) {There are at least two *points*, Every *point lies on* at least two *lines*, There are at least three *lines*}
(Σ2) {There are at least two *points*, Every *point lies on* at least two *lines*, There is at most one *line*}

The relation $R_{\Sigma 1}$ defined by Σ1 is that relation that holds of a triple <*P*, *LO*, *L*> iff *P* is a first-level concept under which at least two objects fall, *L* is a first-level concept under which at least three objects fall, and *LO* is a relation whose domain is *P* and that maps each object falling under *P* to at least two objects falling under *L*.[10] $R_{\Sigma 2}$ is similar except that *L* is a first-level concept under which at most one object falls. There is a straightforward sense in which $R_{\Sigma 1}$ is consistent while $R_{\Sigma 2}$ is not; specifically, $R_{\Sigma 1}$ is consistent in the sense that there is no contradiction involved in the assumption that some *n*-tuple satisfies it, while there is such a contradiction in the case of $R_{\Sigma 2}$. As we'll say, $R_{\Sigma 1}$ is *satisfiable* while $R_{\Sigma 2}$ is not.[11]

If one assumes the consistency of the background theory *B* (in Hilbert's case, a theory of real numbers), then the fact that the members of Σ can be interpreted as theorems of *B* shows, as above, the consistency of Σ. If one assumes further the truth of *B*'s theorems, and particularly the existence of those entities (e.g., sets of real numbers) affirmed by *B*, then the fact that the members of Σ can be interpreted as theorems of *B* shows the existence of an *n*-tuple satisfying R_{Σ}. That is, on this stronger assumption, the interpretation demonstrates the satisfiability of R_{Σ} by exhibiting an *n*-tuple that satisfies it. We will refer below to this *n*-tuple as a "Σ-structure." A Hilbert-style consistency-proof for a set Σ demonstrates (given the stronger assumption of *B*'s accuracy) the existence of a Σ-structure, which more than suffices for the consistency in Hilbert's sense of Σ itself and for the satisfiability of R_{Σ}.[12]

5.4 Frege's Objections

Frege's fundamental objection to Hilbert is that Hilbert's proofs are not about the axioms of geometry. As he puts it in 1914, in a passage in which the phrase "axioms in Euclid's sense" refers to thoughts about e.g. points, lines, and planes rather than about real numbers:

> [Hilbert] is not speaking of axioms in Euclid's sense at all when he discusses their consistency and independence. We could say that the word "axiom," as he uses it, fluctuates from one sense to another without his noticing it. It is true that if we concentrate on the words of one of his axioms, the immediate impression is that we are dealing with an axiom of the Euclidean variety; but the words mislead us, because all the words have a different use from what they have in Euclid.[13]

Recall that for Frege, the axioms of a science are not sentences but thoughts. When we write down a series of sentences in the effort to axiomatize geometry, the role of those sentences is simply that of expressing thoughts; the thoughts themselves are what we're really interested in. Similarly, as we have seen above, Frege takes it that thoughts, not sentences, are the relata of the fundamental logical relations. Questions of consistency and independence are from Frege's point of view always questions about thoughts, and are incoherent if presumed to apply to sentences that express no determinate thoughts. As he puts it in 1903, "There simply cannot be any question of contradiction in the case of senseless sentences."[14] Similarly in 1906:

> We have to distinguish between the external, audible, or visible which is supposed to express a thought, and the thought itself. It seems to me that the usage prevalent in logic, according to which only the former is called a sentence, is preferable. Accordingly, we simply cannot say that one sentence is independent of other sentences; for after all, no one wants to predicate this independence of what is audible or visible. . . .
>
> It must be noted that Mr. Hilbert's independence-proofs simply are not about real axioms, the axioms in the Euclidean sense; for these, surely, are thoughts.[15]

Before turning to the details of Frege's discussion, let's fix some terms. Where Σ is a finite set of geometric sentences of the kind Hilbert deals with, and taking as given a particular one of Hilbert's re-interpretations involving the real numbers:

> – Let T_G be the set of thoughts expressed by Σ when its terms have their ordinary geometric senses. Thus T_G is a set of thoughts about ordinary points, lines, and planes. The members of T_G are the thoughts in whose consistency we are really interested, as Frege sees it. We'll need to refer below also to the conjunction of these thoughts; call this conjunction "$\wedge T_G$."
>
> – Let T_R be the set of thoughts expressed by Σ when its terms take Hilbert's assigned meanings. T_R is therefore, as Frege would put it, a set of thoughts about the real numbers. Similarly, $\wedge T_R$.

> – Let $T_{\exists}$ be the existential thought expressed by the result of conjoining Σ's members and existentially quantifying over the geometric terms. This last thought essentially says that there is some series of concepts, relations, etc. that will satisfy the schematically-understood Σ; it says in short that there is a Σ-structure, i.e. an n-tuple satisfying R_{Σ}.

The important question, from the Fregean point of view, is whether Hilbert's reinterpretation technique suffices to establish the consistency of T_G. His view is that it doesn't.

Frege articulates the central difficulty with Hilbert's approach in two ways. In the first, the focus is on the ambiguity of the geometric terms as Hilbert treats them. In Fregean terms, the difficulty is that Hilbert's focus on the set Σ of reinterpretable sentences leaves it unclear which set of thoughts, if any, he is dealing with. When Hilbert says that the axioms of geometry "express basic facts of our intuition,"[16] he seems to be talking about T_G, the set of thoughts whose consistency is really at issue. But when Hilbert re-interprets Σ's terms via constructions on the reals, he leaves T_G behind and can only be understood to be talking about T_R. And when Hilbert talks about the axioms themselves as having components with no fixed interpretation, he is no longer talking about any particular set of thoughts at all. As Frege puts it in 1914:

> We could say that the word "axiom" as [Hilbert] uses it fluctuates from one sense to another without his noticing it. It is true that if we concentrate on the words of one of his axioms, the immediate impression is that we are dealing with an axiom of the Euclidean variety; but the words mislead us, because all the words have a different use from what they have in Euclid. . . .
>
> [A] sentence that is meant to express an axiom may not contain a new sign. All the terms in it must be known to us. As long as the word "between" remains without a sense, the sentence "If A, B, C, are points on a straight line and B lies between A and C, then B lies between C and A" fails to express a thought.
>
> An axiom, however, is always a true thought.[17]

We can see why this issue would loom large for Frege. From the Fregean point of view, the different thoughts expressible by a given sentence under its geometric and its real-number interpretations may well have radically different logical properties, so that the set of thoughts expressed by a set of sentences under one of its interpretations may be consistent while the set expressed under the other interpretation is inconsistent. Hence the indeterminacy of meanings involved in Hilbert's treatment undermines the idea that one is asking (or answering)

meaningful questions about the consistency and independence of any particular sets of thoughts. For Hilbert, on the other hand, the distinctions Frege insists upon are entirely beside the point. Because Hilbert is concerned with notions of consistency and independence with respect to which the specific meanings of non-logical terms have absolutely no bearing, it makes no difference whether one has in mind a geometric or a real-number interpretation of the non-logical terms; it makes no difference whether one takes the results in question to hold of sentences or of thoughts, and there is no reason (indeed quite the contrary) to require the terms in question to have determinate senses or references.

Frege's second way of framing what he takes to be the central confusion in Hilbert's method concerns the role of partially-interpreted geometric sentences as providing definitions. In response to Frege's complaint that the geometric terms in *FG* are not well-defined, Hilbert replies that each set Σ of sentences provides a definition of its terms.[18] Frege is quick to point out that these "definitions" do not do what one might have hoped for: they do not fix determinate meanings on the terms "point," "line," etc. They do not, to put it in Frege's terms, assign first-level concepts and relations to the geometric predicates.[19] They instead specify general conditions to be met by any of the variety of (n-tuples of) first-level concepts and relations that are to serve as references of those terms, which is to say as Frege puts it that they define second-level concepts or relations. This way of taking the axiom-sentences, namely, as defining multiply-instantiable general conditions, is essentially Hilbert's own point when he explains, as above, that his sets of axiom-sentences each define a "scaffolding or schema of concepts" whose "basic elements can be thought of in any way one likes;" here again, the idea is that the set Σ of sentences defines the multiply-instantiable general condition R_{Σ}. Frege's discussion of the issue is worth quoting at length:

> Now, how do things stand with Mr. Hilbert's definitions? Apparently every single point is an object. From this it follows that the concept of a point (*is a point*) is of the first level, and consequently that all of its characteristics must be of the first level. If we now go through Mr. Hilbert's axioms, considering them as parts of the definition of a point, we find that the characteristics stated in them are not of the first level. That is, they are not properties an object must have in order to be a point. Rather, they are of the second level. Therefore, if any concept is defined by means of them, it can only be a second-level concept. . . . No doubt the relationship of the Euclidean point-concept, which is of the first level, to Mr. Hilbert's concept, which is of the second level, will then have to be expressed by saying that . . . the former falls within the latter. It is then conceivable—in fact probable—that this does not apply to the Euclidean point-concept alone. And this agrees with what

> is said on p. 20 of [*FG*]: "Consider a pair of numbers (*x,y*) of the domain Ω to be a point," etc. . . . We should probably construe the matter thus: The first-level concept *is a pair of numbers of the domain* Ω, just like the Euclidean concept of a point, is supposed to fall within Mr. Hilbert's second level concept. . . .
>
> According to the preceding, Euclidean geometry presents itself as a special case of a more inclusive system which allows for innumerable other special cases—innumerable geometries, if that word is still admissible. And in every one of these geometries there will be a (first-level) concept of a point and all of these concepts will fall within the very same second-level concept. If one wanted to use the word 'point' in each of these geometries, it would become equivocal. To avoid this, we should have to add the name of the geometry, e.g. "point of the A-geometry," "point of the B-geometry," etc. Something similar will hold for the words "straight line" and "plane." And from this point of view, the questions of the consistency of the axioms and of their independence from one another (that is, of the unprovability of certain propositions from certain presuppositions) will require re-examination. One could not simply say "the axiom of parallels," for the different geometries would have distinct axioms of parallels. If the wording of each of these were the same, this would mistakenly have been brought about by the fact that one had simply said, for example, "straight line" instead of "straight line of the A-geometry." This way of talking may veil the difference of the thought-contents, but it certainly cannot remove it.[20]

In short, a set of axiom-sentences in *FG* characterizes a general condition (our R_{Σ}, Hilbert's "scaffolding," Frege's "more inclusive system") satisfiable if at all by series of concepts and relations of various kinds, including those from the realm of geometry and those from the realm of real analysis. The axiom-sentences themselves can be made to express particular sets of thoughts, e.g. T_G and T_R, by assigning particular senses to the geometric terms.

Frege acknowledges that Hilbert's method immediately shows the consistency, i.e. the satisfiability, of R_{Σ}, since the Σ-structure Hilbert constructs satisfies that general condition. Equivalently, by pointing out the truth of each member of T_R, Hilbert successfully establishes the truth and hence the consistency of its existential generalization $T_{\exists}$. Speaking of the method as applied to independence-demonstrations, in which, as we might put it, the point is to show the satisfiability of $R_{(\Sigma \cup \sim\phi)}$, Frege writes to Hilbert as follows:

> From a general logical point of view the case is always the same: you want to show the lack of contradiction of certain determinations. "D is not a consequence of A, B, and C" says the same thing as "The satisfaction of A,

> B and C does not contradict the non-satisfaction of D." "A, B, and C are independent of one another" means "C is not a consequence of A and B; B is not a consequence of A and C; A is not a consequence of B and C." After reducing everything to the same schema in this way, we must ask, What means have we of demonstrating that certain properties, requirements (or whatever else one wants to call them) do not contradict one another? The only means I know is this: to point to an object that has all those properties, to give a case where all those requirements are satisfied. It does not seem possible to demonstrate the lack of contradiction in any other way. If you are merely concerned to demonstrate the mutual independence of axioms, you will have to show that the non-satisfaction of one of these axioms does not contradict the satisfaction of the others (I am here adopting your way of using the word "axiom.") But it will be impossible to give such an example in the domain of elementary Euclidean geometry because all the axioms are true in this domain. By placing yourself in a higher position from which Euclidean geometry appears as a special case of a more comprehensive theoretical structure, you widen your view so as to include examples which make the mutual independence of those axioms evident.[21]

That is to say: taking the higher-level "determinations," "properties," "requirements" etc. expressed by Hilbert's schematically-understood axiom-sentences, or those partially-interpreted sentences themselves, to be the things whose consistency and independence are at issue, Hilbert's T_R stands as the "example" or instance that immediately demonstrates the consistency or independence at issue. By 1906, Frege refers to Hilbert's partially-interpreted axiom-sentences as "pseudo-axioms" and makes the same point, noting that Hilbert's demonstrations show "the independence of the meanings of the parts of pseudo-propositions," i.e. of the higher-level concepts and relations defined by those sentences.

Frege concludes the passage quoted above, addressed to Hilbert, as follows: "Although I am struck here by a doubt, I will not pursue it further here."[22] His reticence doesn't last long. Frege notes repeatedly from 1900 to 1914 that the crucial inference Hilbert seems to be making, from the independence or consistency of "pseudo-propositions" and what they express to the independence and consistency of Euclidean axioms, is fallacious. As he puts it in the second "Foundations of Geometry" essay:

> Mr. Hilbert appears to transfer the independence putatively proved of his pseudo-axioms to the axioms proper, and that without more ado, because he simply fails to notice the difference between them. This would seem to constitute a considerable fallacy. And all mathematicians who think that Mr. Hilbert has proved the independence of the

> real axioms from one another have surely fallen into the same error. They do not see that in proving this independence, Mr. Hilbert is simply not using the word "axiom" in the Euclidean sense.[23]

That is, while the satisfiablity of R_{Σ} is shown by Hilbert's constructions, it is a mistake to suppose that this implies the consistency of T_G.

In response to Korselt,[24] who takes up in 1903 the project of defending Hilbert, Frege writes:

> In saying that modern mathematics no longer designates certain facts of experience with its axioms but at best indicates them, Mr. Korselt brings the axioms of modern mathematics into contrast with those of Euclid; and doubtless we may assume that he counts himself among the modern mathematicians. Clearly, he also counts Mr. Hilbert among them and believes that with this proposition he has hit upon the latter's use of the word "axiom." If this is correct, then it is a gross error to assume that Mr. Hilbert has shown anything at all about the dependence or independence of the Euclidean axioms; or that when he talks about the axiom of parallels it is the Euclidean axiom.[25]

Similarly:

> Mr. Korselt declares that it is not risky to talk of "the theorem *a*" (e.g. the axiom of parallels) if *a* has an identical or similar wording in all geometries, as if the sense did not matter at all. I have given grounds to the contrary; Mr. Korselt ignores them and simply opposes his authority to them. Does this suffice? Well, his manner of talking is indeed practical if it is a question of imagining that what one has proved of axioms taken in the modern sense also holds of them when taken in the Euclidean sense.[26]

Frege's repeated insistence on this point, that Hilbert's procedure is "vitiated by the fact that the sense of the axioms is by no means securely fixed,"[27] can stand as the significant objection Frege takes it to be only if the consistency of T_G really is a separate question from that of T_R, and specifically only if the consistency of T_G is not implied by that of T_R or of its generalization $T_{\exists}$. And indeed, Frege does hold the questions to be distinct in just this way. As he puts it in 1900:

> If a general proposition contains a contradiction, then so does any particular proposition that is contained in it. Thus if the latter is free from contradiction, we can infer that the general proposition is free from contradiction, but not conversely.[28]

The first point is straightforward: if $T_\exists$ is inconsistent, then every instance of it, including T_R and T_G, will be as well. So if either of these special cases has been shown consistent (as indeed one of them has by Hilbert), we can conclude that $T_\exists$ itself is consistent. The second point (the "not conversely") contains the disagreement with Hilbert: the consistency of $T_\exists$ does not imply the consistency of T_G.

Similarly:

> [G]iven that the axioms in special geometries are all special cases of general axioms, one can conclude from lack of contradiction in a special geometry to lack of contradiction in the general case, but not to lack of contradiction in another special case.[29]

As above: from the consistency of T_R we can straightforwardly infer the consistency of $T_\exists$. But—and here again is the source of disagreement with Hilbert—we cannot infer the consistency of T_G from that of T_R.

A similar sentiment appears in Frege's letter to Liebmann in the same year (1900), in which Frege discusses the equivalent question of independence:[30]

> I have reasons for believing that the mutual independence of the axioms of *Euclidean* geometry cannot be proved. Hilbert tries to do it by widening the area so that Euclidean geometry appears as a special case; and in this wider area he can now show lack of contradiction by examples; but only in this wider area; for from lack of contradiction in a more comprehensive area we cannot infer lack of contradiction in a narrower area. . . .[31]

Why does Frege take it that the consistency of T_R and of $T_\exists$ is insufficient to guarantee the consistency of the very-similar T_G? It is worth repeating here that from Hilbert's point of view, the inference in question—i.e. from the consistency of the axioms as interpreted over the reals to their consistency as interpreted over geometric entities—is entirely unproblematic. For as far as Hilbert is concerned, the consistency-question has to do with what these two "special cases" have in common, namely, the overarching structure exemplified by each, and described by the set Σ when its terms are taken schematically. But for Frege, the crucial point is that the consistency-question for T_G does *not* have to do just with the form shared by T_G and T_R, and described by $T_\exists$. It has to do additionally with what distinguishes T_G from T_R, namely, the specific geometric senses that go toward making up T_G.

Frege completes the sentence we've left hanging above as follows:

> . . . from lack of contradiction in a more comprehensive area we cannot infer lack of contradiction in a narrower area; *for contradictions might enter in just because of the restriction*.[32]

You can't infer the consistency of T_G from that of $T_\exists$ or of T_R since the specific subject-matter (the "restriction") introduced by T_G may well give rise to contradiction. A similar sentiment is expressed by Frege in his earlier "On Formal Theories of Arithmetic." Here Frege is discussing different "special cases" of a set of schematically-understood arithmetical rules, one of which is obtained by—as we would put it—interpreting those rules over the natural numbers, and the other by interpreting them over the complex domain. The question is whether the consistency (or the truth) of the former set of thoughts entails the consistency of the latter:

> A proof of non-contradictoriness, then, cannot be given by saying that these rules have been proved as laws for the positive whole numbers and therefore must be without contradiction; for after all, they might conflict with the peculiar properties of the higher numbers, e.g. that of yielding-1 when squared. And in fact, not all rules can be retained. . . . It is therefore evident that in virtue of the peculiar nature of the complex higher numbers there may arise a contradiction where so far as the positive whole numbers are concerned, no contradiction obtains.[33]

I take it that Frege's point is clear: we can introduce a contradiction in the move from T_R to T_G simply in virtue of the move from one subject-matter to another. For as always, for Frege, the question of whether a contradiction follows from a collection of thoughts will depend in part on the subject-matter of those thoughts. It will depend, that is, on the senses had by the non-logical terms of the sentences in question. Though Hilbert's demonstration of the theorem-hood of each member of T_R immediately demonstrates (assuming the consistency of the background theory) the consistency of T_R and hence of $T_\exists$, it is fallacious from Frege's point of view to infer from either of these the consistency of T_G.

The same point can be put in terms of syntactic consistency. Hilbert's demonstration of the theorem-hood of each member of T_R shows, as above and assuming the consistency of the background theory, the syntactic consistency of Σ. No contradiction is derivable via the usual topic-neutral rules from Σ (unless derivable from the background theory). But as we have seen, the syntactic consistency of a set of sentences is no guarantee, from the Fregean point of view, of the consistency of any particular set of thoughts expressible by that set. That Σ is syntactically consistent is compatible with the inconsistency in Frege's sense of the set T_G, which inconsistency would be demonstrable via the derivation of a contradiction from a set Σ' whose members expressed analyzed versions of the thoughts expressed by the members of Σ.

5.5 Consistency and Concepts

Frege's view that syntactic consistency can mask underlying inconsistency at the level of thoughts is closely connected with his view of the logical properties of concepts and relations. As Frege sees it, a given concept can be inconsistent in the sense that a logical contradiction is involved in the supposition that something falls under it.[34] Sometimes this inconsistency will be evident "on the surface," as it were, on inspection of the open sentence used to express it: the predicative phrase ". . . is and is not round" gives such an example. But an important part of Frege's understanding of inconsistency is that such inconsistency can lie well below the surface, in a way that is brought to light only after nontrivial examination of the concept in question. This view underlies Frege's repeated objection to the standard mathematical technique of introducing "new" objects via the two-step procedure of defining a concept under which they are to fall, and then supposing the concept to be instantiated.[35] From Frege's point of view, the first step is unproblematic and requires no demonstration of consistency; inconsistent concepts themselves are perfectly legitimate. The difficulty comes with the second step, which he takes to involve two mistakes. First of all, it's not legitimate simply to suppose without demonstration that a concept is consistent.[36] Second, even if a concept is consistent, this is by itself no guarantee that it's instantiated. As he puts it in *Grundlagen*,

> A concept is still admissible even though its defining characteristics do contain a contradiction: all that we are forbidden to do, is to presuppose that something falls under it. But even if a concept contains no contradiction, we still cannot infer that for that reason something falls under it. If such concepts were not admissible, how could we ever prove that a concept does not contain any contradiction? It is by no means always obvious; it does not follow that because we see no contradiction there is none there, nor does a clear and full definition afford any guarantee against it.[37]

The second point, that the consistency of a concept doesn't imply that anything falls under it, is one that Frege later illustrates via his famous theological example. In the correspondence with Hilbert, he puts it rather gently as follows:

> Suppose we knew that the propositions
>
> 1. A is an intelligent being
> 2. A is omnipresent
> 3. A is omnipotent
>
> together with all their consequences did not contradict one another; could we infer from this that there was an omnipotent, omnipresent, intelligent being? This is not evident to me.[38]

Frege's view here—that instantiation is not implied by consistency—is unquestionably true outside of mathematical contexts. His view that it continues to hold within mathematics is arguably more questionable: one might not-unreasonably argue that in mathematics, the consistency of a collection of defining conditions is all there is to existence. As Hilbert puts it:

> [A]s long as I have been thinking, writing and lecturing on these things, I have been saying the exact reverse: if the arbitrarily given axioms do not contradict one another with all their consequences, then they are true and the things defined by the axioms exist. This is for me the criterion of truth and existence.[39]

More important for our purposes in understanding Frege's conception of logic is the first point made in the *Grundlagen* passage above, namely, that one cannot tell simply by inspection that a concept is consistent. The only way, says Frege, to demonstrate the consistency of a concept is to demonstrate that it is instantiated. In *Grundlagen*, the point is put as follows:

> Strictly, of course, we can only establish that a concept is free from contradiction by first producing something that falls under it. The converse inference is a fallacy. . . .[40]

And in *Grundgesetze*:

> How do we tell that properties are not mutually inconsistent? There seems to be no criterion for this except the occurrence of the properties in question in one and the same object. . . . Or is there perhaps still another way of proving consistency? If there were one, it would be of the highest significance for all mathematicians who credit themselves with a power of creating. And yet hardly anybody seems to concern himself with devising such a type of proof. Why not? Probably people think a proof of consistency superfluous, because any inconsistency would be noticed at once. What a fine thing if it were so! How simple all proofs would then be in their form! The proof of Pythagoras' theorem would go something like this: "Suppose that the square on the hypotenuse were not equal in area to the squares on the other two sides taken together. Then there would be a contradiction between this supposition and the known axioms of geometry. Consequently our supposition is false, and the square on the hypotenuse is exactly equal in area to the squares on the other two sides taken together."[41]

Once again, the crucial fact for Frege is that logical properties and relations can turn importantly on aspects of content that aren't immediately evident even to those working with the thoughts in question. Having understood a collection of

sentences, one is not thereby guaranteed full and perspicuous appreciation of the logical relationships that obtain between the contents of the various sentence-parts. And one is certainly not guaranteed that the syntactically-simple parts of each sentence express logically-simple thought-components or concepts. As Frege says in 1892:

> [S]omething logically simple is no more given us at the outset than most of the chemical elements are; it is reached only by means of scientific work.[42]

Frege's views (1) that a concept can be inconsistent despite the evident syntactic consistency of the predicate expressing it, and (2) that a set of thoughts can be inconsistent despite the syntactic consistency of the set of sentences expressing it, are two sides of the same coin. In each case, the Fregean idea is that conceptual relations, of the kind that can in principle be brought to the surface by good analytic work, can be responsible for logical implications and incompatibilities not yet reflected in the syntactic form of the predicates and sentences in question.

The idea that syntactic simplicity can veil semantic and hence logical complexity is by no means unique to Frege. To begin with, the entirely mundane fact of terminological abbreviation gives ample evidence of this phenomenon: the concept expressed by "prime number" obviously bears an important logical relationship to that expressed by ". . . is evenly divisible by . . .," for the simple reason that the first term is defined in terms of the second. More interesting cases are those in which it's generally agreed that a non-trivial conceptual analysis has brought to light previously-unclear connections; here perhaps the clearest examples involve the conceptual clarifications that lie at the heart of the differential and integral calculus. The idea that what at first appears logically simple can turn out to be surprisingly complex, and hence that apparently-unrelated concepts can be discovered to bear important connections to one another, is not an originally-Fregean doctrine. But what marks Frege off from those who recognize the in-principle possibility of such hidden conceptual connections is his insistence that it is (virtually) always a mistake to suppose that one has already noticed all such connections, or that one can take syntactic simplicity and consistency as a mark of real, underlying logical simplicity and consistency.

5.6 Analysis and Consistency

The difference between Hilbert's and Frege's conceptions of consistency shows up most vividly when we're dealing with a set of sentences some of whose non-logical terms are, as far as Frege is concerned, susceptible to informative conceptual analysis. To return to an example from §1.3 above, Frege takes it that the thought expressed by

(SUM) the sum of two multiples of a number is in its turn a multiple of that number,

which is to say, formalizing a bit, the thought expressed by

(SUMa) $(\forall m)(\forall n)(\forall a)((\text{mult}(m,a) \,\&\, \text{mult}(n,a)) \rightarrow \text{mult}(m+n,a))$

follows logically from the thoughts expressed by

(P1) $(\forall m)(\forall n)(\forall p)((m+n)+p = m+(n+p))$

and

(P2) $(\forall n)(n = n+0)$.

But of course (SUMa) is not derivable via purely-logical rules from the set {(P1), (P2)}, and is clearly, in the sense Hilbert employs in *FG*, independent of that set of sentences. For Frege, the right thing to say about this case is that the relation *x is a multiple of y* can be analyzed in terms of simpler relations and that, when this is done, the resulting analysans-sentence

$$\begin{aligned}(\text{SUM}^{**})(\forall m)(\forall n)(\forall a)((&\forall F[\forall x(Fx \rightarrow F(x+a)) \rightarrow (F(0+a) \rightarrow Fm)] \,\& \\ &\forall F[\forall x(Fx \rightarrow F(x+a)) \rightarrow (F(0+a) \rightarrow Fn)]) \rightarrow \\ &\forall F[\forall x(Fx \rightarrow F(x+a)) \rightarrow (F(0+a) \rightarrow F(m+n))])\end{aligned}$$

is indeed derivable from (P1) and (P2). The original thought expressed by (SUMa) is thereby shown, as Frege sees it, to follow logically from the pair of thoughts expressed by (P1) and (P2). Whenever, as in this case, a conceptual analysis can expose previously-hidden logical complexity, there will be cases in which a sentence ϕ is independent in Hilbert's sense from a set Σ of sentences, while the thought $\tau(\phi)$ follows logically, as Frege sees it, from the set $\tau(\Sigma)$. Similarly, there will in such cases be sets of sentences consistent in Hilbert's sense that express sets of thoughts inconsistent in Frege's. Because, as Frege puts it, contradictions do not always "lie quite open to view,"[43] it can take nontrivial analytic work to discover the contradiction hidden in a set of thoughts. And whenever it does, the Hilbert-consistency of the set of sentences will fail to imply the Frege-consistency of the set of thoughts expressed.

The logical inferences by means of which Hilbert will demonstrate, in *FG*, that a sentence follows logically from others are all acceptable to Frege, in the following sense: if ϕ is derivable in Hilbert's sense from Σ, then $\tau(\phi)$ follows logically in Frege's sense from $\tau(\Sigma)$. Hilbert's derivation will, from Frege's point

of view, express a proof when fully fleshed out. Hence Fregean consistency and independence are stronger than the corresponding Hilbertian relations: if a set of thoughts is consistent in Frege's sense, then any set Σ of sentences whose members express just those thoughts will be consistent in Hilbert's sense. But, as above, not vice-versa: some Hilbert-consistent sets of sentences will express Frege-inconsistent sets of thoughts. Similarly, the independence in Frege's sense of a thought from a set of thoughts implies the corresponding sentential independence in Hilbert's sense, but not vice-versa.

Frege does not claim to have in mind any specific conceptual analyses of geometric concepts that would reveal inconsistencies where Hilbert sees consistency; his point is rather the general one that Hilbert's crucial inference is unwarranted. Frege does claim in his first letter to Hilbert on this topic that in his own investigations into the foundations of geometry, "I thought I could make do with fewer primitive terms."[44] Because these investigations of Frege's have not survived, we have no clear indication of the details of the reduction Frege had in mind here, but the central idea is presumably that of expressing some of the geometric concepts or relations that Hilbert takes as primitive in terms of others. If so, there will certainly be, in the context of geometry, cases like those just sketched in which Σ is consistent in Hilbert's sense while the geometric set of thoughts it expresses is inconsistent in Frege's.

5.7 The 1906 Passage

Frege only once takes up the question of how one might prove independence and, by extension, consistency.[45] His treatment of the issue is very brief and inconclusive, and also difficult to square with the rest of his work. The passage appears in the last part of the 1906 "Foundations of Geometry" essay and comes just after a rehearsal of the problems with Hilbert's method.[46] The interesting part is a sketch of a proposal for demonstrating independence.

The central idea of the proposed method turns on a mapping μ from terms of a language L to terms of L, one that preserves syntactic type[47] and maps "logical" terms to themselves. L is fully interpreted and is "logically perfect" just in the sense that each well-formed sentence expresses a determinate thought. The mapping μ will induce a map μ' from sentences to sentences. Given a set P of premise-sentences all expressing truths, and a conclusion-sentence C expressing a truth, the idea is that we assess the independence of C from P by looking at the set P' of sentences to which the members of P are mapped by μ', and the sentence C' to which C is mapped by μ'. If each member of P' expresses a truth, while C' expresses a falsehood, then our original C (and the thought it expresses) is independent of the set P (and the set of thoughts expressed by its members).[48]

Frege closes his discussion of the test on a cautious note, pointing out two difficulties: first, that of giving the proposed test with more precision, and second, that of distinguishing logical from non-logical vocabulary, as is essential in order to precisely specify the second requirement on the mapping μ. He does not claim here that these difficulties are insurmountable or that they give rise to serious objections to the proposal. His attitude seems to be rather that there is more to be said, and that further investigation is required before the rule can be clearly formulated and applied.

The odd thing about the proposal is not just that it can't work, but that it can't work for exactly the same reason that Hilbert's method can't work: it ignores the contribution of non-logical terms. If we assume that there are no expressive limitations on the language in question, then: given a set Σ of sentences and a sentence s, there will be an interpretation (in Hilbert's sense) on which each member of Σ is true while s is false, if and only if there is a function μ of the kind Frege mentions mapping Σ to Σ', s to s', and such that each member of Σ' is true while s' is false. The only difference between the Hilbert-interpretation I and Frege's function μ is that while I maps each term t to a set or object o, μ maps t to a term t' that refers to o. So as long as the language in question contains names for all of the objects and relations (or their extensions) to which Hilbert has recourse in constructing interpretations, Hilbert's independence-test and the proposed 1906 independence-test will have exactly the same results. In failing to take into account the logical relationships that obtain among the non-logical parts of arguments, the 1906 test will often declare a given sentence "independent" of others when in fact, from Frege's point of view, the thought expressed by that sentence is a logical consequence of the thoughts expressed by those others.

One might take it that the 1906 proposal gives reason to doubt Frege's consistent commitment to the conception of consistency and independence already sketched, on which these relations turn not just on surface-level form but also on the "deeper" relationships revealed by conceptual analysis.[49] Indeed, if Frege really means the 1906 proposal as a fully-general test of independence, then there is a conflict between this proposal and the account offered here of Frege's understanding of the logical relations.

But there are real difficulties with taking the 1906 proposal as representative of Frege's general views about logic and form. Most important is the equivalence just noted between the proposal and Hilbert's own method: if Frege is seriously proposing the "mapping" strategy as a general independence-test applicable to languages containing terms like those found in Hilbert's *FG*, then he is flatly contradicting himself within the space of a single essay. Similarly, Frege's own views about logical implication as these emerge throughout all of his positive work, on which logical implication has everything to do with the contents of such terms as "successor" and "multiple," directly contradict the purely-formal account of implication that seems to lie behind the 1906 proposal.

Frege's hesitant and quickly-abandoned attempt to chart out a method for proving independence displays a number of interesting features of his thought in 1906. This is the only place at which Frege tries to work out the implications for independence-demonstrations of his view of logic itself as entirely general, and hence as a science whose principles apply independently of the subject-matter at hand. As Frege puts it, the test he proposes is an "emanation of the formal nature of logical laws."[50] What he means here is explained as follows:

> [A]s far as logic itself is concerned, each object is as good as any other, and each concept of the first level as good as any other and can be replaced by it; etc.[51]

except that, as Frege quickly adds, the purely logical concepts and relations are not interchangeable in this way. All of the laws of logic to which Frege appeals throughout his work are "formal" in this sense: his axioms and rules of inference are purely general and do not favor any non-logical object or concept over others. This does not of course mean that the logical *relations* are "formal" in this sense: as we have seen, the fact that one thought follows logically from another can turn essentially on the particular objects and concepts involved. As Frege puts it,

> We may say that in a certain sense, every inference is formal in that it proceeds according to a general law of inference; in another sense, every inference is non-formal in that the premises as well as the conclusions have their thought-contents which occur in this particular manner of connection only in that inference.[52]

The apparent tension here—between logical *laws* as formal and logical *relations* as not entirely formal—is no more than apparent. As Frege sees it, the logical relations between thoughts can all be traced back to the formal laws, but only after a thorough analysis of the relevant objects and functions (or the corresponding senses) into simpler constituents. As he puts it, conceptual analysis can "reveal the logical linkage of truths."[53] Thus despite the purely formal nature of the laws of logic, the substitution of one non-logical concept or object for another can make a critical difference with respect to consistency and independence. Hence the formality of the logical laws is not enough to ensure the success either of Frege's 1906 test or of Hilbert's test of independence.

What, then, was Frege thinking in proposing the test? This we will presumably never know, given the sketchiness of his discussion and his failure to return to the issue. Perhaps he briefly confused the formality of the logical laws with

the formality of the logical relations. Perhaps on the other hand he meant the test only for "fully-analyzed" languages, ones whose contents are known to bear no logical relations to one another. If so, the test would be successful where applicable (as would Hilbert's), but it would not have been applicable in any interesting mathematical domain, particularly not in geometry or arithmetic.

However we understand the 1906 proposal, one thing that's clear is that Frege was not happy with it. He never returns to the proposal or to the open questions surrounding it, and by 1910 seems to have concluded that independence cannot be demonstrated. In Frege's notes to Jourdain of 1910, we find:

> The indemonstrability of the axiom of parallels cannot be proved. If we do this apparently, we use the word "axiom" in a sense quite different from that which is handed down to us. Cf. my essays "On the Foundations of Geometry. . . ."[54]

Frege's reader is referred here, presumably, not to the brief positive proposal at the end of the second essay but rather to the discussion constituting the bulk of both essays referenced, i.e., to the criticism of Hilbert. That the second essay contains the outline of a positive solution to the question of independence-demonstrations is not even mentioned. Similarly, the posthumously-published "Logic in Mathematics," written in 1914, contains a rehearsal of the early criticism of Hilbert's independence-proofs, and no mention of the 1906 proposal.

5.8 Ultimate Analyses?

The "gap" between Hilbert-consistency and Frege-consistency arises, again, because syntactically-simple parts of the sentences in question can in principle express what are, from Frege's point of view, logically-complex senses. This raises the question whether in Frege's view there is a "final" level of analysis at which the gap disappears. Is there a system of maximally fine-grained linguistic representations of thoughts such that each Frege-inconsistent set of thoughts is represented in this system by a Hilbert-inconsistent set of sentences? Frege, as far as I can tell, simply does not consider this question, and nothing he says appears to commit him clearly to either an affirmative or a negative answer. Nevertheless, perhaps a few remarks are in order.

In order to demonstrate positive consequence-results, i.e. claims of the form "α is a logical consequence of Γ" for α a thought and Γ a set of thoughts, Frege requires α and Γ to be expressed by a sentence ϕ and set Σ of sentences, respectively, such that ϕ is derivable from Σ. This can, and in the case of the logicist project does, require a certain amount of detailed analysis, since the most familiar sentences used to express α and the members of Γ may not be derivable one from another. In

short, as we've seen in detail above, the demonstration of positive consequence-results will often require conceptual analysis, of the kind Frege takes himself to be providing in *Grundlagen*. What such demonstrations do not require, however, is the claim that the analysis given is in any sense final; there is no need to show, and indeed Frege never attempts to show, that further, more-detailed analysis is impossible. Thus the demonstration of logicism requires no particular view about the existence of a final analysis of thoughts, or of a language in which all logical relations are reflected syntactically.

One reason to suppose that Frege would have denied the existence of such an ultimate level of analysis stems from some of his views about the multiple decomposability of contents.[55] If we take Frege's remarks in *Grundlagen* regarding the decomposition of contents with respect to such sentences as (par) and (dir) as indicative of his general later view regarding the decomposition of thoughts, it would seem that Frege holds that a given thought can be expressed by two sentences neither of which is obtainable from the other by analysis of the simple components of either. An example of this phenomenon would be the expression of a given thought (or of logically-equivalent thoughts) by each of the sentences "The extension of F = the extension of G" and "$(\forall x)(Fx$ iff $Gx)$," and similarly for sentence-pairs having to do with directions and lines, with shapes and orientation, etc. Because neither sentence is obtained from the other by a mere expansion of atomic parts, there is no obvious sense in which one expresses a fuller analysis of the thought than does the other. Frege's acknowledgment of such pairs is a reason to give a negative answer to the question of the existence of a final level of analysis in the sense raised here. The reason is inconclusive, however, since Frege might well have held, if pressed, that such incomparable pairs of sentences are always further reducible to some common sentence expressing a final analysis.

That Frege does not take up in detail the question of the existence of terminal analyses in the context of the debate with Hilbert is not surprising. If we find ourselves dealing with a language whose syntactically simple terms stand only for logical simples, so that no logically-relevant connections of content obtain between these terms, then Hilbert's independence- and consistency-demonstrations will be accurate: a set Σ of sentences of such a language will be syntactically consistent only if the set $\tau(\Sigma)$ is consistent in Frege's sense; similarly for independence. But Frege holds that consistency and independence can't be determined by inspection. As he puts it:

> It is completely wrongheaded to imagine that every contradiction is immediately recognizable; frequently the contradiction lies deeply buried and is only discovered by a lengthy chain of inference.[56]

Hence there is from Frege's point of view never a guarantee that the language in question is in fact "fully analyzed" in the sense in question. And without such

a guarantee, Hilbert-style demonstrations can give no assurance of Fregean independence or consistency.

5.9 Concluding Remarks

Frege is clearly right that, if one understands "consistency" in his own semantically-rich way, then Hilbert's *FG* methodology is not a reliable means of demonstrating consistency. Frege's view that relations of entailment, and hence of inconsistency and dependence, can turn on the semantic content of simple non-logical terms means that consistency and independence in Frege's sense do not follow from the kind of consistency and independence straightforwardly demonstrable via Hilbert's method. And Frege's further view that the discovery of the relevant semantic relations can often require difficult mathematical or other conceptual work means that there is no straightforward way to exploit the Hilbert-style methodology in demonstrating what Frege calls "consistency" and "independence." Hilbert's methods are, as Frege insists, not a reliable means of demonstrating what Frege took to be at issue in the foundations of geometry.

But that Frege is right on this point should not be allowed to obscure his peculiar blindness to the sense in which Hilbert's methods are entirely reliable and fruitful. Indeed, Frege's own formal systems provide just the kind of context best suited to the use of Hilbert-style methods to prove non-derivability results. Suppose we are working with Frege's *Begriffsschrift* system, supplemented in the way Frege suggests with terms for some special science, say geometry or physics.[57] Given a set Σ of sentences of that system, a Hilbert-style reinterpretation *R* on which each member of Σ expresses a truth will guarantee that no contradiction is derivable in that system from Σ (assuming the background consistency of the theory used to construct *R*). Similarly, a Hilbert-style reinterpretation of the members of Σ and of a formula ϕ on which each member of Σ is true while ϕ is false will show (given the same assumption) that ϕ cannot be derived via the Fregean formal rules of deduction from Σ. For Frege's *Begriffsschrift* rules are truth-preserving under any interpretation that assigns meanings in the usual way to the non-logical terms.

Frege would not of course view non-derivability as an indication of consistency or of independence, and he is right to insist on the difference. But the fact remains that non-derivability results are important, that they should have been of interest to Frege, and that Hilbert's method is the best general method for providing them. While it is essential to Frege's view of logic that Hilbert's methods fail to establish consistency or independence in Frege's own rich sense of these notions, nothing in Frege's view stands in the way of acknowledging the success of those methods in establishing non-derivability. And it is notable that,

as far as one can tell, Frege simply never pauses to acknowledge the success in this admittedly limited sense of Hilbert's methods. In partial defense of Frege's silence here, it is worth noting again that while positive derivability-results are of considerable significance to Frege, establishing as they do the provability of a thought from others (assuming the reliability of the formal system), negative such results are of no immediate significance with respect to the question of whether a given thought is logically entailed by others, or of whether a set of thoughts is consistent.

The methods Hilbert exhibits in *Foundations of Geometry* stand today as the core of those contemporary methods that appeal to models in the demonstration of mathematical and logical results. In the next chapter, we survey some of the central ways in which the differences between Frege and Hilbert survive as differences between Frege and more modern approaches to fundamental logical issues.

Notes

1. This material is descended from Blanchette [1996], [2007]. Parts of it are prompted by Hodges [2004], to whom I owe thanks not just for this article but also for helpful correspondence.
2. Hilbert [1900].
3. Frege to H. Liebmann July 29, 1900, [1980] p 90. This sentiment is repeated often throughout the next 14 years, and is never softened by Frege.
4. That is, logical implication in the sense relevant to Hilbert-style consistency proofs is independent of such meanings.
5. Hilbert [1900] §9, p 30 of the English 10th ed. Ω is the field out of which the reinterpretation is constructed.
6. Letter to Frege December 29, 1899, as excerpted by Frege (ellipsis Hilbert's or Frege's), [1980] p 40.
7. This is not to say that Hilbert's rules of derivation in *FG* are syntactically specified. They are instead left implicit; the important point here is that they are straightforwardly syntactically specifiable in now-familiar ways.
8. This requires that there is a canonical order on Σ's terms.
9. Strictly speaking, the sets we should be considering here are:

$$\{\exists x\exists y(Point(x)\ \&\ Point(y)\ \&\ x \neq y),\ \forall x(Point(x) \rightarrow \exists y\exists z(Line(y)\ \&\ Line(z)\ \&\ LiesOn(x,y)\ \&\ LiesOn(x,z)\ \&\ y \neq z)),\ \exists x\exists y\exists z(Line(x)\ \&\ Line(y)\ \&\ Line(z)\ \&\ x \neq y \neq z)\};$$
$$\{\exists x\exists y(Point(x)\ \&\ Point(y)\ \&\ x \neq y),\ \forall x(Point(x) \rightarrow \exists y\exists z(Line(y)\ \&\ Line(z)\ \&\ LiesOn(x,y)\ \&\ LiesOn(x,z)\ \&\ y \neq z)),\ \forall x\forall y((Line(x)\ \&\ Line(y)) \rightarrow x = y)\}$$

10. Equivalently, one can take R_Σ extensionally as a relation satisfied by sets rather than by concepts and relations. Nothing in what follows turns on this distinction.
11. This account of satisfiability is imprecise; for specificity, we would need a crisp account of what it is for the relevant assumption to "involve a contradiction." All that's required for present purposes are the following points: (1) that if R_Σ is satisfied, then it's satisfiable; and (2) that if Σ is syntactically inconsistent, then R_Σ is not satisfiable. In the first-order context, these constraints suffice to fix the relation: R_Σ is satisfiable iff Σ is syntactically consistent (via completeness). In richer languages, further clarification would be required in order to precisely pin down the relation of satisfiability. Nothing in what follows turns on the lack of specificity here.

12. From Hilbert's point of view, the issues of existence and consistency are not separable in the way suggested here, which is to say that the existence of a Σ-structure is not, as Hilbert sees it, a stronger claim than the mere consistency of that existence-statement. For as Hilbert sees it, consistency (in such a context) implies existence. See, e.g., letter to Frege December 29, 1899: "[I]f the arbitrarily given axioms do not contradict one another with all their consequences, then they are true and the things defined by the axioms exist. This is for me the criterion of truth and existence." (Frege [1980] pp 39–40.)
13. [1914] p 267/247.
14. [1903] p 323/277.
15. [1906] p 402/332.
16. [1903] p 322/276.
17. [1914] p 267/247–248. See also the letter to Hilbert January 6, 1900: "It . . . seems to me that there is a logical danger in your speaking of e.g. 'the parallel axiom' as if it was the same thing in every special geometry. Only the wording is the same; the thought content is different in every different geometry." (Frege [1980] p 48.)
18. See, e.g., the letter from Hilbert to Frege December 29, 1899, Frege [1980] pp 38–43.
19. A first-level concept is one under which objects fall. Because chairs, numbers, and geometric points are objects, *chair, number*, and *point* are first-level concepts. Second-level concepts are related to first-level concepts in much the way that first-level concepts are related to objects: to predicate something of a first-level concept is to say that it falls within a second-level concept. For example, to say that *there are F*'s, for *F* a first-level concept, is to say that *F* falls within that second-level concept that one might call *being exemplified*.
20. [1903] pp 374–375/283–284.
21. Letter to Hilbert January 6, 1900; Frege [1980] pp 43–44.
22. Ibid.
23. [1906] p 402/333.
24. Korselt [1903].
25. [1906] p 297/296.
26. [1906] p 307/306.
27. Letter to Liebmann July 29, 1900, Frege [1980] p 91.
28. Letter to Hilbert January 6, 1900, Frege [1980] p 47.
29. Ibid., p 48.
30. For Frege as for us, independence and consistency are two sides of the same coin: ϕ is independent of the set Π iff $\Pi \cup \{\sim\phi\}$ is consistent.
31. Letter to Liebmann July 29, 1900, Frege [1980] p 91.
32. Ibid., emphasis added.
33. [1885] p 102/119.
34. Frege's view here is not as carefully worked out as one might like. In the era of *Grundlagen*, when "concept" refers to the single semantic value of a predicative phrase, the idea that inconsistency attaches to the supposition that such a thing is instantiated is clear enough. But by the time of the mature semantic theory, it is not at all clear that Frege is right to characterize the inconsistency in question as attaching to the *concept* to which the predicative phrase refers rather than to the sense of that phrase. For Frege's view seems to be that the inconsistency in question amounts to the inconsistency of the thought expressed by the existential sentence that involves that phrase. If it is possible on Frege's terms for two sentences $\exists xPx$ and $\exists xQx$ to be such that one expresses an inconsistent (i.e., a self-contradictory) thought while the other doesn't, despite the fact that the predicates refer to the same concept, then concepts are clearly the wrong things to play the role in question. But here we enter a murky area. Much would seem to turn on the identity-criteria for concepts, particularly on the question of whether two predicates related in such a way can yet express the same concept. But for Frege, the idea of identity-criteria for concepts makes, strictly speaking, no sense; identity is a relation reserved for objects. Frege simply says too little about the topic, but I suspect that his talk of inconsistent *concepts* is a holdover from the earlier way of talking, in which concepts play much of the role later played by the senses of predicates.

35. See, e.g., *Grundlagen* (hereafter *Gl*) §§94, 95, 109; [1885] p 103/120; *Grundgesetze* (hereafter *Gg*) 2 §143; [1906c] pp 193–194/179.
36. And to make matters worse, from Frege's point of view, the only way to demonstrate that a concept is consistent is to show that it's instantiated.
37. *Gl* §94.
38. Letter to Hilbert January 6, 1900, [1980] p 47. See also [1903] pp 370–371/280.
39. Hilbert to Frege December 29, 1899, [1980] pp 39–40.
40. *Gl* §95. See also *Gl* §109, in which Frege summarizes one of his objections to the "formalist" theory as follows: ". . . its error lies in taking it as proved that a concept is free from contradiction if no contradiction has revealed itself, and in taking freedom from contradiction in a concept as sufficient guarantee in itself that something falls under it."
41. *Gg* 2 §143.
42. [1892c] p 193/182.
43. *Gg* 2 §144.
44. Frege to Hilbert December 27, 1899, [1980] p 34.
45. For further treatment of this passage, see Ricketts [1998], Tappenden [2000].
46. [1906] §iii, especially pp 426–428/337–339.
47. So that e.g. terms for first-level functions of one argument are mapped to the same, etc.
48. This method is clearly motivated by the very-similar structure of "parallel" proofs provided by duality principles in geometry. For discussion of this point, see Tappenden [2000]. This makes Frege's blindness to Hilbert's method even more peculiar, given Hilbert's claim, in correspondence with Frege, that his own method is motivated by the kinds of parallels provided by duality. See Hilbert to Frege December 29, 1899, Frege [1980] pp 38–43.
49. This is the view of Hodges [2004]. For a reply, see Blanchette [2007].
50. [1906] p 426/337.
51. [1906] pp 427–428/338
52. [1906] p 387/318.
53. [1906] p 303/302.
54. [1980] p 183 note.

Jamie Tappenden [2000] argues that the Jourdain passage is misleading, and that Frege is not here rejecting all independence arguments. As Tappenden sees it, a reasonable reading of Frege is that on which he distinguishes two kinds of independence and corresponding independence-arguments, only one of which he rejects. Independence of the first kind is that in which A is independent of $A_1 \ldots A_n$ if it can be assumed without contradiction that A is false while $A_1 \ldots A_n$ are true. Independence of the second kind is the kind whose demonstration is sketched in 1906. According to Tappenden, the two kinds of independence would be inequivalent in the case of axioms, from Frege's point of view, since for Frege it is impossible for an axiom to be false, and hence impossible for an axiom to be independent of anything in the first sense, though axioms can be independent of others in the second sense.

But this, I think, is not the right way to understand the relevant passages. Frege's objections to Hilbert are not primarily concerned with the unintelligibility of assuming an axiom to be false. The objection is that Hilbert deals with the wrong sets of thoughts, and that the logical relationships between these thoughts can't be used as an indicator of the logical relationships between the thoughts we care about, i.e. the axioms of geometry. Frege's only purpose in pointing out that axioms aren't false is to emphasize that when Hilbert uses an axiom-sentence to express a falsehood about the real numbers, he is very clearly not using that sentence to express an axiom of geometry. Note that (contra Tappenden pp 277–278) Hilbert is not supposing a true thought to be false, or asking what it would be like if it were false, etc. No such counterfactual reasoning is part of Hilbert's argument. Note also that there is, as above, a quick and easy equivalence between Hilbert's proof-strategy and the 1906 sketch. Neither strategy involves assuming an axiom-thought to be false.

As Tappenden points out, the question of whether Frege means in 1910 to reject all independence arguments is separable from the question of the correctness of this specific

understanding of the text. One might, as Tappenden suggests, read Frege's claim that "the indemonstrability of the axiom of parallels cannot be proved" as either (1) making a point about *indemonstrability* in a way which is distinct from independence, or (2) making a limited point about the independence/indemonstrability either of axioms (rather than thoughts in general) or about the axiom of parallels in particular. In either case, the note would not stand as an indication of a rejection of the possibility of independence proofs.

Though the text is, to be sure, brief and inconclusive, it seems to me that neither of these readings is the best way to understand Frege's pronouncement. Frege understands independence in terms of indemonstrability; as sketched in the 1906 essay, a thought is independent of others if we can't demonstrate that thought from those others via correct steps of reasoning. And the "indemonstrability" of the axiom of parallels, in this context, can only mean its indemonstrability from the other axioms of Euclidean geometry. Hence the claim that the indemonstrability of the axiom of parallels can't be demonstrated is effectively a claim that its independence from those other axioms is indemonstrable. Regarding (2), I don't think it likely that Frege would have taken the axiom of parallels, or axioms in general, to differ from the more general class of true thoughts with respect to independence proofs. It's true that Frege took the axiom of parallels to be intimately tied up with the sense of the term "parallel," and that this may well have given that axiom an unusual status for Frege. Nevertheless, the kind of technique tentatively proposed by Frege in 1906, and the technique used by Hilbert in 1900, would have worked just as well (or ill) for the axiom of parallels as for any other axiom of geometry, and indeed for any other thought. Thoughts without such an intimate connection between their constituent senses and their truth might well *be* independent of one another from Frege's point of view, unlike, perhaps, thoughts whose truth-value is more clearly tied to such internal features of meaning; but because neither of the proposed techniques for demonstrating independence is sensitive to such internal facts about constituent senses, neither can be viewed as providing conclusive independence-demonstrations. In short: because a good general method of demonstrating independence would provide a demonstration of the independence of the parallels postulate from the other axioms of geometry, Frege's claim that the latter is impossible is, I take it, a relatively clear indication that he takes there to be no such effective general method.

55. See, e.g., *Gl* §64.
56. [1906c] p 194/179.
57. Frege provides the beginnings of such a supplementation for the case of geometry in e.g. [1879].

6

Frege and Models

6.1 Models and Consistency

Standard logical practice, at the dawn of the 21st century, is to demonstrate consistency via the presentation of models, and indeed in some contexts the term "consistent" has come simply to mean "has a model." The use of models is a natural outgrowth of Hilbert's reinterpretation strategy, and one might well ask whether Frege's criticisms of Hilbert's techniques shed any interesting light on the nature of modern consistency proofs. More generally, one might ask whether those criticisms shed any interesting light on the use of models generally in contemporary logical investigation.[1]

There are, to begin with, four central points of difference between Hilbert's reinterpretation-technique in *Foundations of Geometry* (hereafter *FG*) and the provision of an ordinary model in the sense now typical. First, Hilbert's axioms in *FG* are given roughly in ordinary language: although they involve terms from geometry and are taken as reinterpretable, they are not the sentences of an explicitly-defined formal system. Second, while contemporary models yield truth-values for sentences via a defined *true-on* relation, Hilbert's reinterpretation strategy involves instead simply an implicit understanding of the conditions under which a given sentence is true under a given interpretation. Third, the most common framework within which to provide models these days is that of an axiomatized set theory, while Hilbert used instead a theory of real numbers and constructions thereon whose relevant theorems are simply taken to be evident. And finally, because Hilbert's quantifiers in *FG* are always restricted, he has no need to specify a domain for unrestricted quantification; our own models, on the other hand, regularly involve the specification of such a domain.

Despite these differences, the role of models in demonstrations of consistency and independence is much the same as that of Hilbert-style reinterpretations. To give a model is to demonstrate relative consistency, in the following sense. Suppose we have a formal language L together with a deductive apparatus D for that language, where D specifies the relation of *derivability* between sets of L-formulas and individual L-formulas. We'll assume, as is usual, that derivability

in D turns entirely on formal features of the formulas of L. Let "$\Sigma \vdash_D \phi$" stand for the circumstance that the L-formula ϕ is derivable in D from the set Σ of L-formulas. Suppose also that we have a model-theoretic apparatus M for L, which includes a class of *structures* and a definition of *truth-on-a-structure*, where the latter is a relation holding between L-formulas and structures. Let "$\Sigma \models_M \phi$" stand for the circumstance that ϕ is true on every structure on which each member of Σ is true. Finally, suppose M and D are so related that for every set Σ of L-formulas and every L-formula ϕ: If $\Sigma \vdash_D \phi$, then $\Sigma \models_M \phi$. This last condition, typically called the "soundness" of D (with respect to M), is straightforwardly achieved by choosing as D's axioms only formulas true on every structure, and as its inference-rules only rules that are truth-preserving on every structure.

Given these assumptions, the existence of a structure on which each member of Σ is true, i.e. a *model* of Σ, guarantees the consistency of Σ relative to that of the background set theory in just the way that Hilbert's interpretations guarantee the consistency of sets of sentences relative to that of his background theory of real numbers. Assuming the consistency of that background theory, a model of Σ demonstrates, in short, that no contradiction (i.e., no pair of a sentence and its negation) is derivable from Σ via the use of the deductive system D.

Because deductions in D are insensitive to the meanings (if any) had by the non-logical terms appearing in the sentences of Σ, the question of whether Σ is consistent in the sense demonstrated by a model has entirely to do with the syntactic structure of its sentences, and not with the meanings of its terms. As with Hilbert-style consistency, then, we'll refer to this kind of consistency as "syntactic consistency." The existence of a model of Σ shows that, if the background theory (say, ZFC) is syntactically consistent, then Σ is syntactically consistent.

Frege's sense of consistency is importantly different, for the reasons seen above in the controversy with Hilbert. If a set of thoughts is consistent in Frege's sense, then any set of sentences expressing those thoughts will be syntactically consistent in the present sense, assuming a reasonable deductive system.[2] But the converse is false: that a set of sentences is syntactically consistent does not entail that a given set of thoughts expressed by those sentences is itself consistent in Frege's sense.[3] To choose a straightforward example: consider the set of thoughts {*0 is a natural number, 1 is a natural number, the successor of 0 = the successor of 1*}. Expressing these thoughts in an ordinary language of first-order logic gives us the set of sentences {$N(0)$, $N(1)$, $s(0) = s(1)$}. The set of sentences itself is syntactically consistent, as witnessed by the existence of a model, but of course the set of thoughts is, as Frege sees it, not consistent.

In short, Frege's idea that instances of logical entailment can be revealed by conceptual analysis means that the kind of syntactic consistency demonstrable via models is no guarantee of Fregean consistency.

6.2 Models and Entailment

On a standard and compelling modern picture of mathematical theories, most if not all of the important logical characteristics of theories have to do with their classes of models. Where a *theory* in the modern sense is a set of partially-interpreted sentences, we've already seen two important senses of theory-consistency understood in terms of structures: when a theory has a model, it is consistent both in the sense that the condition it defines is satisfiable and in the sense that no contradiction is syntactically derivable from it. Similarly, we can measure the expressive richness of a theory in terms of the uniformity of its models; here the significant issues include those of categoricity, definability, and other measures of the fineness of grain with respect to which theories divide their classes of models from the whole universe of structures. Finally, once structures themselves become a focus of inquiry, various natural questions arise concerning relations of containment between definable classes of structures, especially that of whether all structures satisfying a given set of sentences also satisfy some further sentence. We have, in short, the issue of model-theoretic "entailment," that of whether every model of a set Σ of sentences also models the sentence ϕ.

We have seen in some detail what Frege thought, and what he should have thought, of the use of models to demonstrate consistency and independence. While he should have recognized the utility of models in demonstrating non-derivability results—i.e. the results we have called above "syntactic consistency" and "syntactic independence"—he was nevertheless right to reject the use of models in demonstrating consistency and independence in his sense, as applied to thoughts. But what about the idea of structures themselves as a focus of interest, and of their use in assessing the expressive richness of theories and in characterizing a version of entailment? Here we take a brief look at how these issues appear from a Fregean point of view.

To begin with, it is worth recalling that Frege takes a theory to be not a set of sentences but a set of thoughts. Hence the idea of a *theory* as characterizing a general condition satisfiable by different structures makes no sense from the Fregean standpoint. Nevertheless, as we've seen, Frege does appreciate the general spirit of the Hilbert-style approach to the characterization of multiply-instantiable structural conditions. The idea of a set Σ of sentences as characterizing a general condition R_Σ instantiable by, essentially, what we would call "structures" is not foreign to Frege.[4]

Similarly, Frege shows clear evidence of appreciating, as one would expect, that a critical feature of a theory's axioms (in his sense of "theory" and of "axiom") is that they pin down with appropriate specificity the structure of their domain. See for example Frege's demonstration at *Grundgesetze* 2 §157 that his accounts of 0 and of successor guarantee that the collection of objects bearing

the ancestral of the successor relation to 0 (namely, the finite cardinals) exhibit exactly the right structure.[5]

Nevertheless, Frege's central focus on thoughts and the relations obtaining between them guarantees that there is a significant gulf between those logical relations naturally characterized in terms of structures and those that fit easily within a Fregean framework. In what follows, we sketch the main lines of this gulf, with an emphasis on the difference between logical entailment as Frege characterizes it and the relation of model-theoretic entailment.

Suppose then that we have a language whose formulas are to be used for expressing thoughts. While Frege generally treats a language as coming along with a single fixed assignment of thoughts to its formulas, comparison with modern practice is facilitated by relaxing this requirement and taking a language to be susceptible to multiple such assignments. The crucial Fregean point to keep fixed is that the logical properties and relations apply to the thoughts themselves, not to the bare formulas.[6]

Call an assignment of thoughts to the formulas of the language a "reading" of that language. We will suppose that the language comes along with general principles governing the assignment of thoughts to appropriate formulas, principles in accordance with which, for example, a sentence of the form $(\phi \,\&\, \psi)$ is assigned the conjunction of the thoughts assigned to ϕ and to ψ, $\sim\phi$ is assigned the negation of the thought assigned to ϕ, and so on. We'll say, in short, that the language comes along with a class of *acceptable* readings. Where τ is a reading of the language, let $\tau(\phi)$ be the thought assigned by τ to the formula ϕ, and let $\tau(\Sigma)$ be the set of thoughts assigned by τ to the set Σ of formulas.

Suppose further, as above, that the language is equipped with a model-theoretic apparatus M. We can now ask whether, for such a language L, it is the case for each of its acceptable readings τ that:

(I) If $\Sigma \models_M \phi$, then $\tau(\Sigma)$ logically implies $\tau(\phi)$

This is the question of whether model-theoretic entailment in L implies logical entailment in the Fregean sense. Before turning directly to this question, we digress briefly to consider the parallel question with respect to derivability. Given a deductive system D, we want to know whether it is the case, for each acceptable reading τ of the language, that:

(II) If $\Sigma \vdash_D \phi$, then $\tau(\Sigma)$ logically implies $\tau(\phi)$

Frege of course held that the answer to this question was affirmative in the case of his own formal languages and deductive systems: the entire point of these systems was that the derivability-relation among sentences was a guarantee of

the logical-entailment relation among thoughts. The guarantee that deductive entailment implies logical entailment in this sense is straightforward: one simply chooses as axioms only sentences that express (under acceptable readings of the language) logical truths, and as inference-rules only those rules that sanction inferential steps of the right kind: ones in which the sentence obtained in such a step expresses a thought logically entailed by those thoughts expressed by the premise-sentences of that step. Say that a deductive system meeting this condition is *reliable*. That is: D is reliable if, for every acceptable reading τ of the language, (II) holds. We now know, as Frege came to discover in 1902, that while the system of *Begriffsschrift* is reliable, that of *Grundgesetze* isn't.

Just as we say that a deductive system D is *reliable* if it satisfies (II) for each acceptable reading, so too we'll say that a model-theoretic relation M satisfying (I) for each acceptable reading is itself *reliable*. The reliability of a model-theoretic system depends on the details of the language, the class of structures, and the definition of *truth-on* in question. Unlike the reliability of deductive systems, there is no uniform method either for ensuring or for demonstrating model-theoretic reliability across languages and systems.[7] In order to lay out some of the issues here, we begin with a few remarks on structures and readings.

Structures and readings are similar in a number of respects, and indeed it is natural to consider either an "interpretation" of a formal language. But there are significant differences between the two. Typically, a structure "interprets" a formal language by specifying a domain (or multiple domains) of quantification, and assigning individuals and sets to the simple individual-, relation-, and function-symbols of the language. A structure also assigns truth-values to formulas, via the language's defined *true-on* relation: to say that a formula ϕ is true-on a structure S is just to say that ϕ bears that defined relation to S. In short, a structure interprets a language by assigning set-theoretic constructions to parts of formulas and truth-values to formulas. A structure does not, however, assign to formulas anything like thoughts.

A reading, on the other hand, "interprets" a language by assigning thoughts to formulas. The truth of a formula under a reading, unlike its truth on a structure, has nothing to do with the defined *true-on* relation. To say that ϕ is true under reading τ is just to say that the thought $\tau(\phi)$ is, in the ordinary sense, true.

Despite their differences, there is a clear sense in which a structure can serve as an extensional version of a reading. While a reading assigns to the sentence "*Fa*" the thought *2 is an even integer*, there is a structure that assigns to "*a*" the number 2 and to "*F*" the set of even integers. In general: say that a reading τ and a structure A for a language *correspond* just in case the domain of A is the same as the range of the first-order variables under τ, A and τ assign the same individual to each individual constant, and A assigns to the function- and relation- terms the extensions of those functions and relations assigned them

by τ.[8] A structure is, essentially, an extensional version of any reading to which it corresponds.

The important parallel between structures and readings is mediated by the *true-on* relation. This relation is standardly defined in such a way that a formula of a given language is *true-on* a structure iff that formula expresses a truth under any corresponding reading. That is, an important characteristic of typical languages and their model-theoretic apparatus is that they satisfy condition (T):[9]

(T) For every formula ϕ of language L, every structure A of L, and every acceptable reading τ of L: If A and τ correspond, then ϕ is true on A iff $\tau(\phi)$ is true.

Typically, the guarantee of (T) is given in part by defining *true-on* so that the semantic role of the logical constants with respect to structures parallels their semantic role with respect to acceptable readings. Thus e.g. if each acceptable reading assigns the meaning *and* to the symbol "&," then the definition of *true-on* will include the stipulation that a formula of the form '$(\phi$ & $\psi)$' is true-on a structure A iff both of ϕ and ψ are as well. The upshot of this principle is that a formula is true on a structure just in case it says something true when read as being "about" the entities, relations, etc. out of which the structure is constructed. This is what makes it reasonable to call the defined relation "true-on." Henceforth, we'll assume that the language L in question together with its model-theoretic apparatus satisfies *T*.

Given a language L together with an intended reading τ and a model-theoretic apparatus, the first question to ask is that of whether that apparatus includes a structure that corresponds to the intended reading. When it does, we are guaranteed that

— If $\Sigma \models_M \phi$, then it's not the case that: each member of $\tau(\Sigma)$ is true and $\tau(\phi)$ is false.

As one might put it: if $\Sigma \models_M \phi$, then the argument $\langle \tau(\Sigma), \tau(\phi) \rangle$ is truth-preserving, as long as M contains a structure corresponding to τ. This for the simple reason that if $\langle \tau(\Sigma), \tau(\phi) \rangle$ is *not* truth-preserving—i.e. if each member of $\tau(\Sigma)$ is true and $\tau(\phi)$ is false—then the structure corresponding to τ is one on which each member of Σ is true and ϕ is false, so that $\Sigma \not\models_M \phi$.

If to *every* acceptable reading there is a corresponding structure in M, then we have the stronger result that

— If $\Sigma \models_M \phi$, then for every acceptable reading τ, the argument $\langle \tau(\Sigma), \tau(\phi) \rangle$ is truth-preserving.

That $\langle\tau(\Sigma), \tau(\phi)\rangle$ is truth-preserving is of course a far cry from $\tau(\Sigma)$'s logically entailing $\tau(\phi)$; any argument with a false premise or true conclusion is truth-preserving. But it's a start.

In the first-order case, if $\Sigma \models_M \phi$, then $\Sigma \vdash_D \phi$ for any ordinary deductive system D; this is just the completeness theorem for first-order logic. If (as is usual) D is reliable, it follows additionally that $\tau(\Sigma)$ logically entails $\tau(\phi)$. So in the ordinary first-order case, not only does $\Sigma \models_M \phi$ imply that for every reading τ the argument $\langle\tau(\Sigma), \tau(\phi)\rangle$ is truth-preserving; it implies as well that all of these arguments are logically valid. The answer to the question regarding (I) is affirmative, if we restrict ourselves to that fragment of a Fregean (or other) formal system later known as "first-order."

It's worth noting that this roundabout demonstration of the reliability of the model-theoretic entailment relation, a demonstration that hinges crucially on the completeness theorem, is one that Frege would not have been able to foresee. The completeness theorem was not just relatively far in the future from his point of view, but also would have been quite surprising. To see this, consider to begin with the case in which $\Sigma = \varnothing$, so that we're concerned with the case in which ϕ is a model-theoretic truth. Even supposing a nice match between structures and readings, the model-theoretic truth of a given sentence would have been a matter, from Frege's point of view, of the truth of each of a given range of thoughts: specifically of all of the thoughts expressible via that sentence under different readings of the language. And that each such thought is in fact true is no reason to conclude that any one of them is provable from purely-logical principles, i.e. that it's a truth of logic. More generally, the fact that each of the arguments expressible by a premise-conclusion pair $\langle\Sigma, \phi\rangle$ has either a false premise or a true conclusion, which is what we can infer from the fact that $\Sigma \models \phi$, is no reason to suppose that, for any such argument, the conclusion is provable via purely-logical principles from the premises.

An interesting example concerns sentences of the form $\exists x(x = a)$. Because Frege (like most of us) bars non-referring terms from his formal language, each sentence of this form will express a truth under any acceptable reading. It will also be true on every structure in a typical model-theoretic apparatus, simply because we routinely require structures to have nonempty domains. Each such sentence will be, in short, a model-theoretic truth in standard settings. But aside from those cases in which "a" is instantiated by a term referring to a Fregean logical object, the thoughts expressed by these sentences will typically be just contingent (or geometrical, or scientific) truths, not truths of pure logic.

There is no deep tension here between Fregean and modern views of logical truth: a reasonable post-Hilbertian reaction to the standard model-theoretic truth of "$\exists x(x = a)$" is to view it as a justifiable glitch in an otherwise-healthy formal apparatus: we gain a good deal of smoothness by requiring nonempty domains while paying the small price of an isolated model-theoretic truth that

doesn't always express just logical truths. But the example does give a good illustration of how model-theoretic truth looks from a Fregean point of view. The model-theoretic truth of "$\exists x(x = a)$" is from that point of view a matter of the truth of a range of existential thoughts:

$$\exists x(x = \varnothing),\ \exists x(x = \text{the Acropolis}),\ \exists x(x = \text{Germany}),\ldots$$

In this case, of course, there is no temptation to infer the logical truth of any one member of the collection from the truth *simpliciter* of them all. The question, with respect to the connection between model-theoretic truth and logical truth, is whether there is ever a good reason to make such an inference.

As we've seen, the answer, from a Fregean point of view, is "yes," in the special case in which we have a complete and reliable deductive system. The case of ordinary first-order logic is a case in point (excluding the existential-sentence glitch just mentioned.).[10] But once we leave the domain in which model-theoretic consequence coincides extensionally with a well-engineered derivability-relation, things begin to look quite different.

Once one leaves the realm of first-order logic (a fragment in which Frege had no special interest) and the completeness theorem falls away, we can no longer establish the reliability of $\models_M$ in the roundabout way given above. There is now, from the Fregean point of view, little to recommend an inference from truth across structures to logical truth, or from model-theoretic entailment to logical entailment. For from Frege's point of view, a logical truth is one that's provable from a handful of self-evidently logical truths, of the kind displayed in *Begriffsschrift* and *Grundgesetze*. Logical entailment, similarly, is a matter of provability via just such self-evident principles. And there is no reason to suppose that a sentence *true-on* every structure will in general express a thought that's so provable, nor that ϕ's truth on every model of Σ will indicate that there is some Fregean proof of the thought that ϕ expresses from the set of thoughts expressed by the members of Σ.

One guarantee that Fregean logical entailment will diverge from model-theoretic entailment in a class of formal systems richer than first-order logic comes from the incompleteness of those systems. If L (e.g. second-order logic with full models) is incomplete, then its model-theoretic relation is not recursively enumerable, and there is no manageable collection of logical principles with respect to which those entailments can be cashed out via proof.

The difference between Fregean and model-theoretic entailment can also be seen by noting the existence of standard model-theoretic truths that don't, from the Fregean perspective, express logical truths. Consider, for example, the language obtained from an ordinary first-order language by adding the quantifier "There exist absolutely infinitely many, "$\exists^{AI}$," where "$\exists^{AI}xFx$" means that there are "too many" F's to form a set.[11] The formula "$\sim(\exists^{AI}x)x = x$" is true on

every ordinary structure, while its natural reading, namely, the proposition *there aren't absolutely infinitely many things*, is presumably false. Truth across structures in this case not only fails to guarantee logical truth; it fails even to guarantee truth. We also see here a clear illustration of what underlies model-theoretic truth: the sentence in question is true on every structure because it expresses a feature (in this case, the feature of having a set-sized universe) that happens to be shared by every structure, at least on paradigmatic ways of defining a model-theoretic apparatus. And this of course doesn't mean that it expresses the kind of thing provable from self-evidently logical first principles. This phenomenon is to be expected once the language in question is rich enough to express general but nontrivial features of structures.[12]

Other interesting examples arise in the language of second-order set theory. Let "ZFC2" abbreviate the conjunction of the second-order ZFC axioms.[13] This sentence, under its ordinary reading, would be widely viewed as expressing a truth.[14] For it describes what we typically suppose to be the case about the (or a) set-theoretic hierarchy. But it is not at all clear that it has a model, since a model would require the existence of an inaccessible cardinal, something not required by either the plausibility or the truth of either ZFC2 or of any other widely-accepted principles of set theory. If there are no inaccessible cardinals, then ZFC2 expresses, under its "intended" reading, a truth, but has no model. In this case, the negation of ZFC2 is true on every model, but expresses a falsehood.

These examples help to illustrate the gap that we post-Fregeans should recognize between logical entailment as Frege understands it and the model-theoretic entailment relation of modern, post-Fregean systems. But they don't tell us much about whether Frege would or should have noticed such a gap with respect to the precursors of those systems. For the important gap illustrated in the examples just given has to do with the existence of readings to which no structure corresponds, and this gap, based as it is on the view of structures as restricted to well-behaved, non-universal domains, is not part of the general picture of structures which Frege would have been aware of in the first decade or so of the twentieth century.

The less-anachronistic question to ask about Frege, avoiding talk of well-regimented structures, is that of whether he would or should have had any sympathy with the general idea that preservation of truth *across interpretations*, with "interpretations" left deliberately vague, is a good indicator of logical entailment. Similarly, should a Fregean take truth across interpretations to be good indicator of logical truth? Here again, if we take this question as restricted to the first-order fragment of a formal language, one for which there is a reliable deductive system that we know to be complete with respect to the collection of interpretations we have in mind, then the answer is mixed. While the extension of the model-theoretic entailment relation for such a restricted language will be reliable (as above), this is guaranteed not by any a priori general connection

between truth across interpretations and logical entailment in the Fregean sense but by the weakness of the language and the accompanying completeness theorem. And otherwise, barring other such informative restrictions of the question, the answer is "no." Given a suitably well-analyzed language (one whose primitive non-logical terms bear no important logical relations to one another), it will generally be true that the sentences expressing Fregean logical truths will be true across interpretations. But the converse can't be relied upon, from the Fregean point of view. For again, to know of a sentence that it is true across a range of interpretations is simply to know that each of a particular range of thoughts is in fact true; it's not to know that any one of those thoughts can be proven in the appropriate way from self-evidently logical premises.

6.3 Implications

When a *theory* is understood as a partially-interpreted set of sentences, and hence as in large part a means of defining general structural features of mathematical (or other) domains, models are all-important. To know that a theory has a model is to know that it meets two minimal criteria of acceptability: the structural condition it defines is satisfiable, and the set of sentences itself is syntactically consistent. A theory that has "too many" models, e.g. ones exhibiting radically different structural properties, is relatively uninteresting, while one all of whose models are isomorphic, or all of whose models of given cardinalities are isomorphic, for example, is for that reason of real interest. Finally, the relation of model-theoretic entailment, cashing out as it does connections among structural conditions definable via one's theories, is an important relation in its own right.

What we have seen so far is that none of this holds on the Fregean conception of logic. Where Σ is a set of sentences, the important logical questions from Frege's point of view all have to do with the set $\tau(\Sigma)$ of thoughts expressed by the members of Σ. And the existence and characteristics of *models* of Σ are not closely connected with the logical properties and relations of $\tau(\Sigma)$ and its members. While a model of Σ demonstrates, from the Fregean point of view, the syntactic consistency of that set of sentences, it does nothing to guarantee the consistency of $\tau(\Sigma)$. Similarly, the fact that Σ's axiom-sentences are syntactically independent of one another tells us nothing about whether $\tau(\Sigma)$'s axioms (the thoughts expressed by Σ's axiom-sentences) are independent in Frege's sense. The fact that Σ has finite models tells us nothing about whether $\tau(\Sigma)$ is satisfiable in a finite domain. (Consider the theory $\{N(0), \forall x(Nx \rightarrow \exists y Syx)\}$, where τ assigns to "N" and "S" *cardinal number* and *successor*, respectively.) The fact that Σ is too weak to carve out an important or mathematically-interesting class of structures tells us nothing about whether $\tau(\Sigma)$ is a theory of central mathematical interest. And so on. The fact that much of the logical interest of a theory

$\tau(\Sigma)$ turns for Frege on what's expressed by the non-logical terms in Σ means that Σ's class of models, though interesting in its own right, does not offer a way of getting at the logically-interesting features of $\tau(\Sigma)$ itself. The same point can be seen by noting that from Frege's perspective, the theory $\tau(\Sigma)$, or an adequate analysis thereof, can also be expressed by a set Σ' of sentences with a radically-different syntactic structure and hence with a very different class of models.

Most importantly, perhaps, as we've seen, there is no straightforward connection between model-theoretic entailment and logical entailment as Frege understands it. Here the failures occur in both directions. Instances in which $\Sigma \not\models \phi$ although $\tau(\Sigma)$ logically implies $\tau(\phi)$ are straightforward and appear whenever the logical entailment in question is due to relations of containment, incompatibility, etc. obtaining between the contents of simple non-logical terms. In cases of this kind, a further analysis of those contents will yield syntactically-richer Σ' and ϕ' expressing the original thoughts (or their analyses) and such that for some reliable deductive system D, $\Sigma' \vdash_D \phi'$.

Failures in the other direction, i.e. cases in which $\Sigma \models \phi$ but $\tau(\Sigma)$ does not logically entail $\tau(\phi)$, appear whenever (1) every structure in the model-theoretic apparatus either fails to satisfy the condition defined by Σ or satisfies the condition defined by ϕ, and (2) this fact is due not to logic but to uniformities across the chosen collection of structures. As we've seen, these failures occur only outside of the realm of ordinary first-order logic.

Examples can be multiplied. The central point here is just this: when, as in Frege's view, the logical relations of interest obtain between thoughts as opposed to structures, and particularly when those relations turn in part on features of those thoughts that are not all reflected in the syntactic structure of the sentences expressing them, then the logical relations are not definable in terms of models.

This means, among other things, that the question of whether a formal system is *complete* in the sense of Gödel [1930] is one that is not of great interest from a Fregean point of view. It is of course an interesting technical question, given a model-theoretic apparatus, whether every syntactically-consistent set of sentences has a model, i.e. of whether the relation of deductive entailment contains that of model-theoretic entailment. But the idea that such a situation demonstrates the "completeness" of the deductive system in the sense of containing each of the derivations it in some sense "ought" to contain is not one that makes sense from the Fregean point of view. For again, given the focus on thoughts as opposed to structures, and the failure from Frege's point of view of logical entailment to track syntactic features of sentences, the relation of model-theoretic entailment is not the relation that formal derivations were, from Frege's point of view, intended to reveal. For the same reason, the fact that a formal system, like that of Frege's *Begriffsschrift*, fails to be "complete" in this sense is no reason, from the Fregean point of view, to think that it is in any sense defective.

6.4 Summing Up

The use of contemporary structures in the demonstration of consistency and their use in the demonstration of logical entailment are both problematic from the Fregean point of view, but for quite different reasons. The first is due to the fact that consistency in Frege's sense is sensitive to the semantic values of terms, and not just to the syntactic structure of the sentences in question. The second turns on the fact that an argument can fail to be logically valid in Frege's sense, i.e. can fail to be such that the conclusion is provable in a purely logical way from its premises, despite the fact that each of a range of similar arguments (those expressible by the same sentences under readings that correspond with structures for the language) happen all to be truth-preserving.

A natural question to ask is that of whether the first failing can be straightforwardly met by simply insisting on the use of languages none of whose simple terms express senses or references that bear nontrivial logical relations to one another. One might, for example, insist on a language all of whose simple terms express logically-simple objects and functions.

The reply to this suggestion, as noted in Section 5.8 above, is that while such a fully-analyzed language will indeed provide a context in which the existence of a model implies Fregean consistency, the insistence that one's language meet this condition can never be a part of general logical or mathematical practice. For as we've seen, Frege takes it that the relations of entailment and inconsistency that hold between the contents of one's terms are often discoverable only after considerable effort, and indeed may take centuries of mathematical work to uncover. This is why, as Frege sees it, the question of the consistency of a concept can never be settled except by exhibiting something that instantiates it: the fact that a concept or function appears consistent on the surface, i.e. from its syntactic representation, is no guarantee against contradictions that lie below the surface.

> [I]t does not follow that because we see no contradiction there is none there, nor does a clear and full definition afford any guarantee against it.[15]

Similarly, the question of whether a term's content is simple, in the sense of admitting of no further conceptual analysis, is one that as Frege understands it often admits of no easy answer.[16]

We are generally, Frege seems to think, in a situation in which the logical relations borne to one another by the concepts, functions, and objects with which we deal are at least partly known to us, but in which we can't suppose that we know all there is to know on this score. We are not, as it were, logically omniscient. And the fact that we don't generally know that we've hit "rock bottom" with respect to conceptual analysis means that we cannot suppose that our apparently-simple and apparently-independent contents are either simple or

independent. The fact that conceptual analysis is nontrivial is the other side of the same coin: that it can take hard work to break a thought and its components down into simpler components means that it can take hard work to discover previously-unnoticed relations of logical entailment and inconsistency.

Hence while Fregean consistency does follow from syntactic consistency in the right kind of language, the supposition that one's language is of this kind is, from the Fregean point of view, a dangerous one to make. The price for making it mistakenly is that one will declare independence when the thoughts in question bear relations of entailment to one another, and will declare consistency with respect to sets of thoughts that are inconsistent.

The second failing is even less remediable than the first: the Fregean emphasis on thoughts as opposed to structures means that the richer one's language is, i.e. the greater the extent to which one can express nontrivial features of the structures in question, the further the relation of model-theoretic entailment will be from the relation of logical entailment as Frege understands it.

Finally: as we've seen, a noteworthy feature of Frege's approach to logic is that he gives no general characterization of the principles or truths of logic. He simply exhibits a small handful of what he takes to be self-evidently logical truths and inference-rules without claiming to have given an exhaustive list, and counts as logical truths any thoughts provable from those primitive logical truths via the primitive rules of inference. It's not just that Frege doesn't characterize the truths of logic as those (things expressible by) sentences true across interpretations; he gives, more to the present point, no characterization at all of this body of truths aside from its open-ended characterization in terms of proof. Similarly for entailment.

This means that the adequacy of a formal deductive system cannot be demonstrated by proving that its relation of derivability contains, or is contained in, an independently-characterized sentential relation. There is a sense, then, in which Fregean metatheory is and must be quite different from modern metatheory. This is not to say, however, that metatheory in general is foreign to the Fregean perspective. Indeed, quite the contrary: some straightforward metatheoretical issues are central to the Fregean project. The sense in which Frege does metatheory, and the implications of this for our understanding of his conception of logic and of formal systems, is the topic of Chapter 7.

Notes

1. Parts of this chapter come from Blanchette [2000].
2. For example, one satisfying the condition of *reliability*; see Section 6.2.
3. For again, a set of thoughts is consistent in Frege's sense only if no contradiction is logically entailed by that set, where the relation of logical entailment is understood in Frege's rich way, a way that turns not just on syntactic form but also on the meanings of the terms in question.

4. See Sections 5.4 and 5.5 in this volume. Modern practice is to take the instantiating n-tuples extensionally, in Hilbert's way; Frege's way of thinking of them is in terms of concepts and relations (rather than the things falling under these). This difference is worth glossing over for present purposes.
5. For an enlightening discussion of this demonstration, see Heck [1993]. See also Section 7.4 in this volume.
6. Strictly speaking, the important point as before is that, from the Fregean point of view, the sentences have logical properties and bear logical relations to one another only in virtue of what they express; whether this is understood in Frege's way in terms of thoughts is not critical. See Chapter 8.
7. For a fuller discussion of this issue, see Blanchette [2000].
8. This presupposes that τ assigns a range to the first-order variables and assigns individuals, functions, and relations (perhaps via senses that represent them) to the individual constants, function-terms, and relation-terms, respectively. Readings that do not do so (e.g., readings that simply assign propositions directly to wffs) will not be said to correspond to models.
9. This constraint of course does not apply to all of those things commonly called languages; it does not apply, for example, to languages whose model-theoretic apparatus includes Boolean-valued models. But it does apply to all those languages that have a coherent notion of model-theoretic *truth*, and for which the question of the truth of (I) can reasonably arise.
10. One might wonder how the existential-sentence counterexample is possible given the argument for the reliability of $\models_M$ from completeness and deductive reliability given above. The answer is that the deductive system D is not, in this case, strictly speaking reliable (if it's complete), as is shown by the derivability of the existential sentences.
11. This quantifier is discussed by Vann McGee [1992]. For the origin of the quantifier, see Cantor [1899].
12. For extended discussion of this and related issues, see Etchemendy [1990].
13. The second-order axioms are those obtained from the first-order by replacing axiom-schemes (replacement, separation) by their universally-quantified second-order versions. See Shapiro [1991], esp. Chs. 4, 5.
14. See McGee [1992] pp 273, 292.
15. *Grundlagen* §93.
16. See [1892c] p 193/182.

7

Metatheory

Metatheory does not play the central role in Frege's work that it plays in modern logic. Frege never formulates anything like a completeness theorem, or even a completeness question, for his formal systems, and never as far as one can tell devotes an essay or even a lecture to an examination of the adequacy of those systems. What little metatheory Frege does is carried out in passing, via brief arguments presented here and there during the course of introducing his systems. No systematic examination of standard metatheoretical issues appears in Frege's writings, and some of the metatheoretic questions that we find most natural are not raised at all.

Part of the explanation of Frege's silence on what are now regarded as standard metatheoretical questions is simply his historical location: those techniques and results that make e.g. modern completeness and soundness questions appear entirely natural were just not part of the mathematical canon while Frege was working. But there are also some important philosophical differences between Frege and his successors that come to a head just here, over the kinds of metatheoretical questions that make sense, and over those that count as critically important, from the different points of view in question. The purpose of this chapter is to examine the question of how metatheory in general, and some particular metatheoretical questions that form a central part of modern logical theory, look from a Fregean point of view.

7.1 Frege's Metatheory

All of the metatheoretical reflections in Frege's work have to do with establishing that his formal system is adequate for the purposes to which he intends to put it. Because his central purpose is to demonstrate the purely-logical grounds of arithmetic, it's essential that each derivation from the empty set of premises make abundantly clear the purely-logical status of the thought expressed by its conclusion. Frege also takes it that the framework he constructs will be applicable, on the introduction of new primitive terminology, to other mathematical and non-mathematical sciences. "I am confident," says Frege, "that my ideography can be successfully

used wherever a special value must be placed on the validity of proofs. . . ."[1] For general applications, it is essential that a derivation from premise-sentences make unquestionably clear that the thought expressed by its conclusion-sentence follows logically from the thoughts expressed by those premise-sentences.

We'll repeat here the definition of *reliability* introduced in Chapter 6, in this case as applied to a language that, like Frege's, has a single canonical reading: A deductive system *D* for such a language is *reliable* iff, for each of its sentences ϕ and each of its sets Σ of sentences:

> If ϕ is derivable in *D* from Σ, then the thought expressed by ϕ follows logically from the thoughts expressed by the members of Σ.

It is crucial for Frege's purposes that his system be reliable, and also essential that its reliability be relatively transparent. The fundamental difficulty revealed by Russell's paradox was, again, that the enriched system of *Grundgesetze* was not in fact reliable.

The systems of *Begriffsschrift* and *Grundgesetze*, which we'll call "*B*" and "*G*" respectively, each include a clearly-specified language, a collection of axiom-sentences, and a handful of rules for deriving sentences one from another. In order that the systems be obviously reliable, it is essential that each axiom-sentence express a thought that isn't just a logical truth, but one that is so simple and self-evident that its purely-logical status is beyond question. Similarly, it is essential that the derivation-rules sanction only the most self-evidently logical inferences, that they be applied correctly, and that the introduction of defined terms causes no problems:

> Because there are no gaps in the chains of inference, every "axiom," every "assumption," "hypothesis," or whatever you wish to call it, upon which a proof is based is brought to light; and in this way we gain a basis upon which to judge the epistemological nature of the law that is proved. Of course the pronouncement is often made that arithmetic is merely a more highly developed logic; yet that remains disputable so long as transitions occur in the proofs that are not made according to acknowledged laws of logic, but seem rather to be based upon something known by intuition. Only if these transitions are split up into logically simple steps can we be persuaded that the root of the matter is logic alone. I have drawn together everything that can facilitate a judgment as to whether the chains of inference are cohesive and the buttresses solid. If anyone should find anything defective, he must be able to state precisely where, according to him, the error lies: in the Basic Laws, in the Definitions, in the Rules, or in the application of the Rules at a definite point. If we find everything in order, then we have accurate knowledge of the grounds upon which each individual theorem is based.[2]

The possibility in principle of finding something "defective" in the deductive system is the possibility of finding that an axiom ("basic law") or rule doesn't state or instantiate an "acknowledged law of logic" or that the introduction of a defined term presupposes either some falsehood or some truth of a non-logical kind.

Frege is of course convinced that neither *B* nor *G* has such defects. At the time of *Begriffsschrift*, with its relatively-simple language and semantics, Frege has little to say about definitions other than the standard constraint that they be entirely stipulative and eliminable in principle.[3] He is more explicit about the correctness of his axioms and derivation-rule. In the early sections of *Begriffsschrift*, for example, Frege introduces the conditional sign and its associated derivation-rule, *modus ponens*, as follows:

> §5 If *A* and *B* stand for contents that can become judgments . . ., there are the following four possibilities:
>
> 1. *A* is affirmed and *B* is affirmed;
> 2. *A* is affirmed and *B* is denied;
> 3. *A* is denied and *B* is affirmed;
> 4. *A* is denied and *B* is denied.
>
> Now
>
> $$\vdash (B \rightarrow A)$$
>
> stands for the judgment that *the third of these possibilities does not take place, but one of the three others does*.
>
> . . .
>
> §6. The definition given in §5 *makes it apparent that* from the two judgments
>
> $$\vdash (B \rightarrow A) \textbf{ and } \vdash B$$
>
> the new judgment
>
> $$\vdash A$$
>
> follows. Of the four cases enumerated above, the third is excluded by
>
> $$\vdash (B \rightarrow A)$$
>
> and the second and fourth by
>
> $$\vdash B,$$

so that only the first remains.[4]

Similarly for the axioms. Formula (1), Frege's first axiom, is introduced as follows:

$$\vdash \big(a \to (b \to a)\big)$$

says "The case in which *a* is denied, *b* is affirmed, and *a* is affirmed is excluded." *This is evident*, since a cannot at the same time be denied and affirmed.[5]

The justification of formula (2) is lengthier:

$$\vdash \big(c \to (b \to a)\big) \;\to\; \big((c \to b) \to (c \to a)\big)$$

means "The case in which

$$(c \to b) \;\to\; (c \to a)$$

is denied and

$$\big(c \to (b \to a)\big)$$

is affirmed does not take place." But

$$\big(c \to (b \to a)\big)$$

means the circumstance that the case in which *a* is denied, *b* is affirmed, and *c* is affirmed is excluded. The denial of

$$(c \to b) \;\to\; (c \to a)$$

says that $(c \to a)$ is denied and $(c \to b)$ is affirmed. But the denial of $(c \to a)$ means that *a* is denied and *c* is affirmed. Thus the denial of

$$(c \to b) \;\to\; (c \to a)$$

means that *a* is denied, *c* is affirmed, and $(c \to b)$ is affirmed. But the affirmation of $(c \to b)$ and that of *c* entails the affirmation of *b*. That is why the denial of

$$(c \to b) \;\to\; (c \to a)$$

has as a consequence the denial of *a* and the affirmation of *b* and *c*. Precisely this case is excluded by the affirmation of

$$\big(c \to (b \to a)\big).$$

Thus the case in which

$$(c \to b) \;\to\; (c \to a)$$

is denied and

$$\big(c \to (b \to a)\big)$$

is affirmed cannot take place, and that is what the judgment

$$\vdash \big(c \to (b \to a)\big) \to \big((c \to b) \to (c \to a)\big)$$

asserts.[6]

In each of these cases, Frege's task is the simple one of pointing out that, given what the signs mean, "it is apparent" that the inference-rule is reliable, or "it is evident" that instances of the axiom-sentence scheme express only truths, given just the fundamental logical principle against self-contradiction. Each of the remaining axiom-sentence schemes is accompanied by discussion of this kind, sometimes simply pointing out in German what a formula of the kind in question says, letting its self-evidence speak for itself, and sometimes adding to this a bit of discussion designed to bring the reader to see that what's expressed is indeed an "acknowledged law" of logic, i.e. that it can be seen to be true just on the basis of rudimentary logical reasoning.

It will be important in what follows to note that in these passages of *Begriffsschrift* Frege is not attempting the impossible task of "justifying" the fundamental logical principles in virtue of which contents or thoughts follow from or contradict one another. There is of course no way to do this, since any such attempted justification would presuppose the very principles it attempts to justify. As Frege says with respect to the law of non-contradiction, "If other persons presume to acknowledge and doubt a law in the same breath, it seems to me an attempt to jump out of one's own skin against which I can do no more than urgently warn them."[7] Nothing like (noncircular) argument, or justification, can be given in cases like these, in which what's at issue is a fundamental principle of logic. Frege's purpose in these early pages of *Begriffsschrift* is simply to justify the inclusion of specific axiom-sentences and sentential derivation-rules in his formal system. In each of the brief

discussions to this effect, he relies on his audience to make a number of simple logical inferences and to recognize as correct a handful of basic logical principles. The argument for the reliability of *B* simply takes these fundamental principles and entailment-patterns for granted and uses them to establish the reliability of the sentential derivation-rules and axiom-sentences under discussion.[8]

By the time of *Grundgesetze* and the increased complexity of its language, its semantics, and its deductive system, Frege recognizes that establishing the adequacy of the formal system requires a more extended argument regarding the formulas themselves. In order that axioms always express logical truths, and that derivation-rules applied to sentences expressing truths always yield sentences expressing truths, it is crucial that each sentence of the formal language express a determinate thought with exactly one truth-value. This requires among other things that, as Frege puts it, "every correctly-formed name is to denote something, a principle that is essential for full rigor."[9] In keeping with the importance of this requirement, Frege gives a series of careful treatments of terms and their denotations. Some examples of this approach as found in *Grundgesetze* 1 are as follows:

In §10, as discussed in Chapter 3 above, Frege removes an indeterminacy regarding certain sentences involving course-of-value names, an indeterminacy left by his earlier (§3) introduction of those names. The indeterminacy is removed by a two-step procedure. First, Frege proves that the system as defined to this point can be consistently extended by adding to it any of a number of arbitrary stipulations identifying a particular course of values with the True and another with the False.[10] Second, he gives just such a stipulation. This is the stipulation that the True is to be identified with the course of values of the concept (or of any concept) under which it and it alone falls; similarly for the False.

In §11, Frege introduces the function $\backslash\xi$, his version of the definite article. He notes that the definite article of ordinary language (schematically, "the *F*") carries with it the dual "logical danger" of forming ambiguous names (in the case in which there are two *F*'s) and of forming empty names (in the case in which there are no *F*'s).[11] To ward off this danger, Frege assigns an arbitrary but determinate denotation to his newly-introduced terms in all such cases. He closes the section by pointing out that this strategy works, i.e. that "[t]his danger about the definite article is here completely circumvented. . . ." Neither empty nor ambiguous names can be formed via the new operator, given the rules in terms of which Frege introduces it.

§29 contains an extended discussion of the conditions under which singular terms and function-names denote. §§30 and 31 provide a detailed argument to the effect that correctly-formed proper names and names of functions all denote.[12]

In §32, Frege notes that each name of the formalism has not just a denotation but also a sense, and that in the case of sentences, this sense is a thought.[13] Having set out in §33 the principles of correct definition and applied these in §34 to the

definition of the symbol "$\frown$," Frege notes in §35 that the definition is indeed a good one in the sense that, as is essential for the "correctness of . . . proofs," the defined function-term has a denotation.

With respect to rules of derivation, Frege's discussions are similar to those of *Begriffsschrift*. The principle that he calls the "interchangeability of subcomponents" is introduced in §12 via the following line of reasoning:

> "$\Delta \to (\Lambda \to \Phi)$" . . . denotes the False if Δ is the True and $(\Lambda \to \Phi)$ is not the True, i.e. in this case if $(\Lambda \to \Phi)$ is the False. But this is the case if and only if Λ is the True and Φ is not the True. Accordingly,
>
> $$\Delta \to \big(\Lambda \to \Phi\big)$$
>
> is the False if Δ and Λ are the True while Φ is not the True; in all other cases it is the True. *From this there follows* the interchangeability of Λ and Δ;
>
> $$\Delta \to \big(\Lambda \to \Phi\big)$$
>
> is the same truth-value as
>
> $$\Lambda \to \big(\Delta \to \Phi\big).$$[14]

Modus ponens is introduced as follows:

> From the propositions "$\vdash (\Delta \to \Gamma)$" and "$\vdash \Delta$" we may infer "$\vdash \Gamma$"; for if Γ were not the True, then since Δ is the True, $(\Delta \to \Gamma)$ would be the False.[15]

Similarly for the next rule:

> The following method of inference is a little more complicated. From the two propositions "$\vdash (\Delta \to \Gamma)$" and "$\vdash (\Phi \to \Delta)$" we may infer the proposition "$\vdash (\Phi \to \Gamma)$." For $(\Phi \to \Gamma)$ is the False only if Φ is the True and Γ is not the True. But if Φ is the True then Δ too must be the True, for otherwise $(\Phi \to \Delta)$ would be the False. But if Δ is the True then if Γ were not the True then $(\Delta \to \Gamma)$ would be the False. Hence the case of $(\Phi \to \Gamma)$'s being the False cannot arise; and $(\Phi \to \Gamma)$ is the True.[16]

Axioms are justified similarly, in much the same way as in *Begriffsschrift*, though with more explicit appeal to truth-values. Basic Law I, which is Formula (1) of *Begriffsschrift*, is introduced as follows:

$$\big(\Gamma \to \big(\Delta \to \Gamma\big)\big)$$

could be the False only if both Γ and Δ were the true while Γ was not the True. This is impossible; therefore

$$\vdash \left(a \to \left(b \to a\right)\right)^{17}$$

Similarly for Basic Law IV:

> —Δ and —~Δ are always different, and always truth-values. Now, since —Γ is in any case always a truth-value, it must coincide either with —Δ or with —~Δ. It follows from this that (~(—Γ = —~Δ) → (—Γ = —Δ)) is always the True; for it could be the False only if ~(—Γ = —~Δ) were the True (i.e., if —Γ = —~Δ were the False) and —Γ = —Δ were not the True (i.e. were the False). In other words, (~(—Γ = —~Δ) → (—Γ = —Δ)) could be the False only if both —Γ = —Δ and —Γ = —~Δ were the False, which, as we just saw, is not possible. Therefore,

$$\left(\sim\left(—a = — \sim b\right) \to \left(—a = —b\right)\right).$$

And so on for the remaining axioms: Frege points out in each case that, given what the signs mean, it is impossible for an instance of an axiom to refer to the False.

These metatheoretic passages consist essentially in (1) arguments to the effect that the syntactic formation-rules and meaning-stipulations together guarantee that every well-formed piece of language has an appropriate reference and that well-formed sentences express determinate thoughts; and (2) arguments, or in simple cases mere pointings-out, that each axiom and rule of derivation is reliable. The arguments and discussions take up a very small portion of the two formal works and are by no means at the center of Frege's attention. These passages do, however, suffice for Frege's metatheoretic purposes: they show (or would have shown, if they hadn't contained errors) that Frege's formal systems were reliable, and hence that Frege's means of demonstrating his logicist thesis, and of setting up a framework for rigorous logical demonstrations in general, was a good one. From the vantage point of one who takes the formal system as an important tool, and not as an independently-interesting topic for investigation in its own right, this is just what one should expect to see as metatheory.

7.2 Universalism and Metatheory

7.2.1 The Issue

The account just given, on which Frege provides careful arguments regarding the adequacy, in various senses, of his formal systems, is not uncontroversial. A number of commentators on Frege have held that Frege's conception of logic makes it impossible or incoherent for him to engage in this kind of metatheoretic reasoning.

In particular, the view in question is that Frege conceives of logic as "universal" in a sense that makes the whole idea of metatheory incoherent.

Dreben and van Heijenoort, for example, hold that:

> [N]either in the tradition in logic that stemmed from Frege through Russell and Whitehead, that is, logicism, nor in the tradition that stemmed from Boole through Peirce and Schröder, that is, algebra of logic, could the question of the completeness of a formal system arise.
>
> For Frege, and then for Russell and Whitehead, logic was universal: within each explicit formulation of logic all deductive reasoning, including all of classical analysis and much of Cantorian set theory, was to be formalized. Hence not only was pure quantification theory never at the center of their attention, but metasystematic questions as such, for example the question of completeness, could not be meaningfully raised. . . . We have no vantage point from which we can survey a given formalism as a whole, let alone look at logic whole.[18]

Similarly, Goldfarb holds that:

> If the system constitutes the universal logical language, then there can be no external standpoint from which one may view and discuss the system. Metasystematic considerations are illegitimate rather than simply undesirable.[19]

The spatial metaphor, according to which there is no "vantage point," no "external standpoint," no "perspective" from which to carry out metatheoretic reasoning, is pervasive in this line of interpretation. Metatheory, on this account, requires one in some sense to "stand outside" the formal system about which one is reasoning, but, the view continues, Frege's understanding of logic or of his system as "universal" entails that it is impossible to do so.

There are a number of things one might mean by the claim that logic or a system of logic is "universal," some of which do, and some of which do not, have much to do with metatheory. We'll begin with two straightforward ways of interpreting the term and its associated line of reasoning, if only to clear the ground for the more interesting and more complex versions.

One sense in which one might be a "universalist" about logic is simply to hold that logic as a whole, i.e. that collection of principles underlying all correct inference, is universal in the sense that it applies everywhere and (hence) that it serves as the grounds of all justification and explanation. As Ricketts puts it:

> Any explanation will draw on the principles of logic. In this way, logic, the maximally general science, provides a framework that embraces every science. . . . Indeed, because of logic's maximal generality, as Frege

> understands justification and explanation, no other science can have justificatory or explanatory relevance to logic.[20]

Frege is, to be sure, a universalist in this sense. From the Fregean perspective, logic applies everywhere, and the relation of logical entailment is essential to the justificatory force of all explanations and lines of argument. From this form of universalism, it follows immediately that one can never engage in non-circular justifications of the fundamental principles of logic or of, in Dreben and van Heijenoort's [1986] term, "logic as a whole." Any such justification would presuppose the very principles it seeks to justify.

But this restriction has nothing to do with metatheory. Metatheory is never an attempt to justify logic as a whole; it is always an attempt to evaluate particular formal systems, i.e. particular codifications of those universal logical principles. At issue in metatheoretic investigations are the virtues of the codification, not of the underlying principles being codified.[21]

A second sense of "universalism," more in line with the first quotation above, is that in which a formal system S is taken to be universal in virtue of being applicable everywhere, to all areas of inquiry. Frege was certainly a universalist in this sense; his view was that his formal systems, suitably modified by adding vocabulary as required, could serve as frameworks for presenting proofs not just about arithmetic but about any area of discourse in which rigor of proof was at a premium. But this form of universalism, again, brings with it no difficulties for metatheory: that a system is (intended to be) universally applicable in this sense is no barrier to our asking meaningful questions about it, e.g. about whether it does in fact have the expressive and deductive resources required for universal applicability, about whether it is reliable, and so on.

We turn now to conceptions of universalism that are arguably more promising in providing difficulties for metatheory.

The barrier to metatheory arises not when one takes it simply that logic in general is universal or when one takes it that one's formal system is universally applicable, but when one holds the considerably stronger thesis about that formal system that its derivations offer the *only* way of presenting compelling or scientifically-acceptable arguments. The idea that Frege holds such an exclusivist position about his formal system is taken to be supported by two lines of thought in Frege. The first is that the inadequacies of ordinary language, including primarily the ambiguity and unclarity of its terms, make ordinary language an unsatisfactory vehicle for the expression of compelling or scientifically-acceptable arguments. The remedy for this inadequacy is taken to be the expression of all such arguments via derivations in the formal system. The second line of thought is that compelling or scientifically-acceptable arguments require expression within a system of "unified science." For Frege, as Ricketts reads him:

> The justification of knowledge is its systematization. Frege understands systematization through the lens of his logical work as the economical axiomatic formalization of a branch of knowledge within the framework provided by the *Begriffsschrift*.[22]

Similarly,

> Proofs receive definitive expression only inside the framework the *Begriffsschrift* provides for all sciences.[23]

The idea that a formal system *S* is universal in this sense—i.e. that it formalizes all of scientific inquiry in such a way that no scientifically-acceptable arguments can be given except via derivations in *S*—causes problems for metatheory in two ways.

The first difficulty is one of circularity: if the only way to present legitimate or scientifically-compelling arguments is via derivations in *S*, then all attempts to give such arguments presuppose the reliability of *S*. Hence, to try to argue in a legitimate or scientifically-compelling way for *S*'s reliability would be, by the universalist's lights, to reason in a very small and vicious circle.[24]

It is worth noting that this circularity-argument does not rule out all metatheory. The metatheoretic claims one might make about a formal system can be divided into two camps as follows. The first contains just those claims whose truth is presupposed when we treat derivations within the system as expressing justification-conferring proofs. Included here are most obviously the reliability of the system, its consistency, the truth-preservation of its derivations, and various forms of soundness. Into the second camp fall those claims about formal systems that are not presupposed in making such justificatory appeal to its derivations, i.e. those metatheoretic claims about a system whose falsehood regarding that system does not undermine its reliability. Included here, for example, are various claims of comprehensiveness for the system and completeness in the modern sense. That a system fails to include *all* of the derivations of a certain class does not mean that the derivations it does include are in any way faulty. The exclusivist position regarding a formal system *S* immediately and obviously entails that those metatheoretic claims about *S* that fall into the first camp—the "reliability" camp—cannot be non-circularly demonstrated. But there is no such quick route from exclusivism about *S* to the indemonstrability of those second-camp claims (those from the "comprehensiveness" camp) about *S*. That we would have to presuppose the *reliability* of *S* in order to demonstrate its comprehensiveness or its completeness—in the same way that *S*'s reliability would be presupposed in order to demonstrate anything at all, on this line of argument—does not make such demonstrations circular.[25]

The second difficulty posed for metatheory by this form of universality is that of semantic paradox. In order to obtain from the system of *Grundgesetze* a formal system *U* sufficient not just for e.g. physics and astronomy, but also for semantics, one would need to add primitive terms for, among other things, the fundamental properties of interest in metatheory. And we know that, when the subject of the metatheoretic investigations is *U* itself, this can't be done. If *U* is consistent, then it must lack some of the resources (e.g., a well-behaved negation operator, referring terms for its own formulas, generous rules of sentence-formation, a well-behaved truth-predicate, and so on) that one might reasonably take to be essential for metatheory. If Frege held that the only scientifically-acceptable investigations are those that can be carried out in a formal system for unified science, then he held a view from which very little metatheory can count as scientifically acceptable.

Does Frege hold such a form of universalism? In particular, does he hold either (a) that scientifically-acceptable arguments can only be presented via derivations in a system U of unified science, a system that contains the primitive vocabulary for all of the (coherent) sciences, or (b) the weaker claim that scientifically-acceptable arguments can only be presented via derivations in a rigorous formal system that incorporates essentially the logical resources of Frege's own *G* or *B*? Note that the global problems for metatheory arising from semantic paradox are difficulties only on the assumption of the strong thesis (a), since the weaker requirement that the metatheory of e.g. Frege's system *G* be carried out in an appropriately-rigorous formal system is easily met by pursuing that metatheory in the system *G*' obtained from *G* by adding a truth-predicate for *G*. The problem cannot of course be circumvented if there's only one acceptable system *U* within which both arithmetic and semantics must be carried out, as proposed by (a). The problem of circularity outlined above arises for metatheory on the assumption of either (a) or (b).

As to (a): Frege clearly held that the framework provided by his formal system was to have been applicable quite broadly. He remarks in 1882 that:

> I did not wish to present an abstract logic in formulas, but to express a content through written symbols in a more precise and perspicuous way than is possible with words. In fact, I wished to produce, not a mere *calculus ratiocinator*, but a *lingua characteristica* in the Leibnizian sense.[26]

Whether he means here that his system will be like a Leibnizian *lingua characteristica* simply in the sense of expressing determinate contents via its formulas, or in the further sense of (ideally) sufficing for the expression of all scientific thought, is not made clear in this text. Frege does talk in *Begriffsschrift* of possible later expansions of his system to include e.g. geometry and physics, and in [1882] discusses future expansion as follows:

> I have attempted to supplement the formula language of arithmetic with symbols for the logical relations in order to produce—at first just for arithmetic—a conceptual notation of the kind I have presented as desirable. This does not rule out the application of my symbols to other fields. The logical relations occur everywhere, and the symbols for particular contents can be so chosen that they fit the framework of the conceptual notation.[27]

The possibility of adding new primitive symbols to the language, and of thereby gaining a richer system applicable to special sciences, is one that Frege seems generally committed to. As he says, "the logical relations occur everywhere," and hence the perspicuous representation of proofs via *Begriffsschrift*-style derivations can be of use well outside the domain of arithmetic.[28]

Frege does not, however, claim that there will be an eventual über-system containing the resources necessary to do all of science. More to the present point, he does not claim that scientific acceptability must await such a final expansion of his system to all of science. Quite the contrary. Frege's own system *G* is far from being universal in the relevant sense, containing only extremely minimal expressive resources; but of course Frege takes derivations in *G* to be very much up to the standards of scientific acceptability. (Indeed, as we'll see below, he takes them to go well beyond those standards.) It may well be that "universalism" in the strong sense of (a) was never meant by those advocating a universalist reading of Frege; in any case, this reading is a non-starter.

Considerably more interesting is the idea that the inadequacies of ordinary language prompt Frege to hold that scientifically-acceptable arguments can only be given via derivations in either *G* or some expansion thereof. Frege does indeed claim that ordinary language is the source of a certain amount of logical confusion. He holds that an argument's presentation via a derivation in his formal system is a good way of warding off some of the difficulties that beset many arguments, difficulties having to do with the assumption of hidden premises, the use of ambiguous terms, and so on. But that these difficulties beset some non-formalized arguments is never cited by Frege as a reason to take all non-formalized arguments to be in any sense suspect or deficient. He does not claim that sufficiently-careful arguments carried out in natural language are to be regarded as second-rate, and does not claim that the presentation of arguments within his *Begriffsschrift* system is the only way to provide compelling or scientifically-acceptable grounds for conclusions.

The kinds of arguments ruled out by the universalism in question—namely, justification-conferring arguments carried out in a language other than that of a Fregean formal system—include not just arguments concerning the metatheory of that system but also all arguments in the history of mathematics (and elsewhere) prior to Frege's own work. The bold thesis that none of these arguments is sufficient to confer justification on its conclusion is one that does not sit well with Frege's

unremarkable attitude toward the mathematics of his predecessors—namely, that while some of it is problematic in very specific ways, mathematical proof has generally served to establish its conclusions. Commenting on Dedekind's considerably less-rigorous proof-procedure, Frege notes:

> [H]is procedure may have been the most appropriate for his purpose. . . . The length of a proof ought not to be measured by the yard. It is easy to make a proof look short on paper by skipping over many intermediate links in the chain of inference and merely indicating large parts of it. Generally people are satisfied if every step in the proof is evidently correct, and this is permissible if one merely wishes to be persuaded that the proposition to be proved is true.[29]

His own extraordinarily-careful procedure, as Frege sees it, goes well beyond what's required simply to establish the truth of the propositions proven. His logicist purposes, which include not merely the demonstration of the truth of arithmetical claims, but a clear exhibition of the fundamental grounds of each of the (typically already-known) premises, requires that "considerably higher demands must be placed on the conduct of proof than is customary in arithmetic."[30] The rigor imposed by expressing one's proofs as derivations in a Frege-style formal system is, in short, considerably more demanding than is required of arguments sufficient for establishing the truth of their conclusions, even within the demanding field of mathematics.

Frege does not claim that the work of previous mathematicians must be re-cast as formal *Begriffsschrift*-style derivations in order to be persuasive, and does not claim that careful arguments couched in ordinary language or in ordinary-cum-mathematical language are unscientific or otherwise illegitimate. This is as one should expect of an author whose careful philosophical arguments in *Grundlagen*, e.g. to the effect that numbers are objects, that statements of number are assertions about concepts, that arithmetic is not about ideas, and so on, are intended to be taken seriously. Similarly for Frege's metatheoretical arguments in *Begriffsschrift* and *Grundgesetze*.[31]

7.2.2 Internal Tensions

A final difficulty for the exclusivist account of Frege's understanding of his formal system is that it implies an important incoherence in Frege's views. As Ricketts [1985] notes, there is a clear conflict between the exclusivist thesis and Frege's general understanding of our appreciation and application of fundamental logical principles. As Ricketts sees it, the inconsistency is simply a part of Frege's views; one might, on the other hand, view it as a further reason not to read Frege in the proposed exclusivist way.

According to Ricketts [1985], Frege identifies logic with his own formal system. To return to the spatial metaphor introduced by Dreben and van Heijenoort [1986], the idea is that there is no logic "outside" of the system, which is to say that if we reason (in a way that is up to scientific standards), then we are invoking the principles explicitly laid down in Frege's construction of the formal system. To recognize in a sufficiently careful way that a conclusion follows from premises is to appeal—and to appeal *only*—to those axioms and rules that Frege has written down as rules for his formal system. As above, the fact that all knowledge of logical implications must, in order to be scientifically acceptable, appeal to those axioms and rules is, on this view, the heart of the problem for metatheory.

As Ricketts notes, this account is inconsistent with Frege's views about how we recognize simple and basic relations of logical implication between thoughts. There is a strain, as he puts it, between:

> the role for logic defined by Frege's underlying view of judgment and his identification of logic as the maximally general science. For Frege, the ability to reason, to determine what follows from or contradicts what, is intertwined with the understanding of language, the grasp of thoughts.[32]

If you don't understand that the thought that it's raining is inconsistent with the thought that it isn't raining, then you have failed to grasp at least one of these thoughts. The tension arises because, on this account, Frege holds that:

> The logician aims to state principles that enable us to make these determinations in a more regular, less haphazard fashion.[33]

That is to say, the logician writes down statements of logical principles such that the careful reasoner can reach the conclusions of arguments from premises just by appeal to those very statements. But if the recognition of basic relations of logical entailment and inconsistency is a necessary part of the grasp of those thoughts, then it cannot be the case that the recognition of these relations is mediated by the recognition of the truth of some further statements, those written down by the logician. Indeed, the problem is not just that the view of the logician's job as presented in the last sentence quoted is inconsistent with Frege's own view of our means of apprehending logical relations; the view itself is incoherent. For, as Ricketts notes, "no *statement* can formulate a logical principle."[34] The difficulty is that to grasp a basic relation of implication or of inconsistency between some thoughts that one is currently entertaining is not to grasp some *further* thought. Hence the recognition of the most basic relations of implication and of inconsistency cannot be something that one does by simply following the instructions given by the statements of a particular formal system.

Following Ricketts [1985], it's helpful to note that this is just the difficulty pointed out by Lewis Carroll's fable of Achilles and the tortoise. In that story, the tortoise affirms a conditional and its antecedent but refuses to affirm the consequent of that conditional. The interesting point is that he also happily affirms a general statement of the rule *modus ponens*. But adding the statement of this rule to the stock of premises to which he assents makes no difference: the tortoise's position does not *thereby* become inconsistent. The tortoise is obliged to infer the consequent of the conditional not because of his affirmation of the additional premise but because of the logical connection between a conditional, its antecedent, and its consequent. If the tortoise is not already reasoning in accordance with this and other principles that *link* statements one to another, the addition of further statements to his set of accepted premises will never change that. In short, the recognition of simple and basic entailments and inconsistencies between thoughts cannot be a matter of recognizing the truth of further thoughts.[35]

Ricketts's [1985] view is that Frege is just mistaken about this important point, and that the rule-statements of his formal system are intended to serve as the court of appeal for careful reasoning. We are, on this account, supposed to be able to justify conclusions on the basis of premises just by adding to those premises an appropriate stock of Fregean rule-statements and axioms, independently of any underlying capacity to recognize logical relations between thoughts. This is the sense in which logical principles are "identified with" the rules of the formal system. And of course this cannot make sense: if we are not already capable of recognizing logical connections between thoughts, then we won't gain that capacity by adding new thoughts to our collection. Perhaps more to the point: if a conclusion does not in fact follow logically from premises, then adding to that stock of premises a statement of a logical rule will not bring about an entailment.

The picture of a formal system's rule-statements as forming the basis of scientific reasoning, to the exclusion of our ordinary capacity to recognize relations of entailment and inconsistency, is not just incoherent and inconsistent with Frege's views about our apprehension of basic entailment-relations; it is also not what Frege says. Frege does indeed take it that the invention of his formal system will bring a needed degree of precision and clarity to the giving of proofs. A derivation that follows the rules laid down in *Begriffsschrift* is sure to express a gap-free proof, one that demonstrates with a high degree of certainty that its conclusion follows logically from its premises. That is the whole point of the system. But the reason we can be sure that the derivations are reliable in this way is that we can see, via our ordinary and carefully-employed capacity to recognize logical entailments and inconsistencies, that the axioms and derivation-rules are well chosen. As above: the justification for including the sentential derivation rule *modus ponens* among the rules of the system rests on (1) the fact

that "→" expresses the material conditional and (2) the fact that a material conditional and its antecedent together entail its consequent. This last fact is one that Frege relies on his audience to recognize, and it is only because we can and do recognize this fundamental pattern of entailment that we can see that the sentential rule is a reasonable one to include in a system whose derivations are supposed to express good proofs. The benefit of a formal system, from the Fregean point of view, is that by its use we are forced to make explicit, by making very small and clear steps of inference, exactly which inferential principles we appeal to in the course of a proof; it is not to somehow allow us to avoid the use of those principles.

To sum up the discussion to this point: Frege is clearly a universalist in a number of ways. He holds that the fundamental principles of logic apply everywhere; there is no area of inquiry in which these principles somehow "fail." He holds that his system of logic is applicable everywhere, given suitable expansions of primitive vocabulary. He arguably holds that every valid argument can in principle be shown to be valid via a derivation in (some expansion of) his formal system. But he does not hold that logical inferences can only be scientifically justified by appeal to the rules of his formal system. He takes it that not just his mathematical and scientific predecessors, but also he himself, can offer scientifically-acceptable arguments by expressing the thoughts in question in ordinary mathematical German and by employing the ordinary fundamental principles of reasoning that underlie all good argument. This is just what we see him doing in his early mathematical work, in the crucial argumentative passages of *Grundlagen*, and in the metatheoretic reasoning of *Begriffsschrift* and *Grundgesetze*.[36]

7.3 Soundness, Completeness, and Consistency

There is nothing about Frege's conception of logic which rules out metatheory, where "metatheory" is understood as the systematic examination of the adequacy, in various senses, of a formal system. As we have seen, Frege engages in a certain limited amount of metatheoretic reasoning himself regarding his systems *B* and *G*, primarily to the effect that those systems are reliable.

But the amenability of Frege's general approach to broadly metatheoretic reflection does not mean that there is any natural affinity between Frege's position and what one might think of as the core of modern metatheoretic reasoning. As we've seen, Frege's conception of logical truth and logical entailment as applying to thoughts rather than to partially-interpreted sentences, and his view of logical truth and entailment in terms of provability via a core of primitive and self-evident logical principles, means that he cannot take model-theoretic truth or model-theoretic entailment as the touchstone of the logical relations. Hence, the completeness or the incompleteness of a formal

system of logic doesn't have the same significance from the Fregean point of view as it has from some post-Hilbertian viewpoints.[37] There is no sense, from a Fregean point of view, in which a system's completeness is by itself a guarantee that the system includes all of the derivations it was intended to include, or in which a system's incompleteness is a guarantee that it "misses" some logical entailments.

This doesn't mean that a Fregean must be unable to make sense of, to prove, or otherwise to appreciate the implications of modern soundness, completeness, and incompleteness results. Given a language with a deductive and a model-theoretic apparatus, it is a straightforward question whether the extension of the model-theoretic entailment relation for that language includes or is included by the derivability-relation for that language, and a question whose demonstration requires nothing problematic from the Fregean point of view. It's just that the modern point of view from which the adequacy of a deductive system is a matter of its agreement with the model-theoretic entailment relation is one that a Fregean has no reason to share. In particular, the completeness of the first-order fragment of the systems *B* and *G*, and the incompleteness of the full systems, are results that a Fregean can find significant and interesting; but they are results that provide, by themselves, reason neither to recommend nor to fault those systems.

A modern soundness result is of more interest from a Fregean point of view. A system failing soundness is one whose derivability-relation outstrips its model-theoretic entailment relation, so that for some Σ and ϕ, ϕ is derivable from Σ despite the fact that some structure verifies each member of Σ while falsifying ϕ. If readings and structures for the language correspond in the way described in Chapter 6, then there will in such a case be an acceptable reading τ of the language such that each member of $\tau(\Sigma)$ is true and $\tau(\phi)$ is false. Assuming that the language is a Fregean one, in the sense that it comes with an intended assignment τ_I of thoughts to sentences, one can ask in this case two questions: (1) whether the truth of each member of $\tau(\Sigma)$ and the falsehood of $\tau(\phi)$ imply that the intended set of thoughts $\tau_I(\Sigma)$ fails to logically entail the intended thought $\tau_I(\phi)$; and (2) whether there is in such a case something problematic about the derivability of ϕ from Σ. As to the first: because for Frege the logical-entailment relation between thoughts has to do not just with the syntactic form of the sentences expressing them but also with the contents of the non-logical terms, there is no way to answer question (1) in full generality. In some cases, the answer will be affirmative, and in some cases negative: that the thoughts expressed by Σ under one reading don't logically entail the thought expressed by ϕ under that reading doesn't imply that the thoughts expressed by Σ under a different reading fail to logically entail the thought expressed by ϕ under that second reading. But things are clearer with respect to the second question. Though the non-logical terms in Σ and in ϕ may have semantic values in virtue of which $\tau_I(\Sigma)$ logically entails $\tau_I(\phi)$, it will nevertheless be simply a mistake if the deductive system,

which itself appeals only to the syntactic form of the sentences, enables one to derive ϕ from Σ. Because the deductive system is not sensitive to the contents of simple non-logical terms, but only to the syntax of sentences, a condition on its reliability is that it contain the derivation of ϕ from Σ only if under every acceptable reading τ, $\tau(\Sigma)$ logically entails $\tau(\phi)$. And this condition entails that under no acceptable reading is it the case that each member of $\tau(\Sigma)$ is true while $\tau(\phi)$ is false. In short, if for each structure there is a corresponding acceptable reading, as is typically the case, then the reliability of the deductive system requires that it satisfy a modern, model-theoretic soundness result. While soundness, relying as it does on a notion of "truth on a structure" that has little independent interest for Frege, would not have been a natural result for him to investigate, it is nonetheless a result in which a Fregean has a legitimate interest, given its status as a necessary condition for reliability.[38]

Frege does of course intend that the formal system *G* will provide a comprehensive set of axioms for arithmetic, and hence for various arithmetical sub-theories. Consider for example the theory of the natural numbers. Let A_G be that collection of sentences of *G*'s language that express truths about the natural numbers. Because of his realism, Frege will take A_G to be complete, in the sense of including for each sentence of the relevant fragment L' of the language either that sentence or its negation. So the Fregean requirement is simply straightforward theory-completeness as restricted to L': For every sentence ϕ of L', either $G \vdash \phi$ or $G \vdash \sim\phi$.[39]

The modern result most closely related to Frege's interests is a proof of consistency for a formal system. Consistency, i.e. the fact that no formula and its negation are both theorems of the system, is an obvious necessary condition for reliability in any language containing negation. A consistency-demonstration wouldn't have been a natural exercise for Frege to have included in *Begriffsschrift* or *Grundgesetze*, just because the stronger reliability result was, Frege thought, easily seen. But for a system not antecedently known to be reliable, a proof of consistency would count, from the Fregean point of view, as an important first step in ensuring reliability. The modern proofs that the first-order fragment of Frege's system is consistent, and indeed that some interesting second-order fragments are as well, are of real significance from the Fregean point of view.[40]

7.4 Categoricity

Frege's defense of logicism was to have consisted in the proof not, of course, of all of the infinitely many truths of arithmetic, but of a handful of "basic" such truths from which, as he intended, the rest would in turn be provable. A particularly interesting collection of these basic truths is the collection governing the arithmetic of the finite cardinals, i.e. of the collection of objects bearing the ancestral of the successor-relation to 0. Here the basic truths are the claims that

the successor relation is a function, that no member of that collection bears the ancestral of the successor relation to itself, and that every member of that collection has a successor.

As Richard Heck points out, Frege establishes in his proof of *Grundgesetze*'s theorem 263 that, as one might put it, these fundamental truths provide a categorical characterization of the natural numbers. That is, Frege establishes that, in Heck's words:

> . . . [For any relation $Q\xi\eta$ and any object a] if $Q\xi\eta$ is functional, if the G's are the members of the Q-series beginning with a, if no member of the Q-series beginning with a follows itself in the Q-series, and if each G is related, by $Q\xi\eta$ to some object, then the number of G's is *Endlos* because the G's, ordered by $[<_Q]$, are *isomorphic to the natural numbers*, ordered by $[<_{PRED}]$, that is, by less-than.[41]

This crucial result is one that might reasonably be called "metatheoretical," and it is worth asking to what extent it is so from a Fregean point of view.

To begin with, the usual modern way of framing such a categoricity result is as a claim about structures: taking some of the terms of the language as uninterpreted, we ask, as did Hilbert in the case of geometry, which coordinated assignments of meanings or extensions to those terms will satisfy the set of sentences. The result in this case is that any pair of assignments to "0" and to "*Pred*" that satisfies the basic sentences will be respectively an object o and a relation R such that the collection of objects to which o bears the ancestral of R is isomorphic (under $<_R$) to the natural numbers (under $<$).

Frege of course doesn't put it quite this way, given his emphasis not on the satisfaction of partially-interpreted sentences by structures but on the thoughts and the relations expressed by (parts of) fully-interpreted sentences. From Frege's point of view, the point is that of the truth, and indeed of the provability in G, of a single universally-quantified thought, one that says essentially that, for any pair $<o, R>$ that satisfies the relation defined by the basic sentences, the closure of o under R is isomorphic (under $<_R$) to the natural numbers (under $<$).

We can put the difference schematically as follows. Let $BL(0, Pred)$ be the conjunction of the basic sentences that Frege derives concerning the natural numbers. The terms "0" and "*Pred*" are shorthand for longer expressions referring to (Frege's versions of) zero and the predecessor relation. $BL(0, Pred)$ expresses a true thought about the natural numbers, one that's equivalent to the conjunction of the second-level Dedekind–Peano axioms. A modern version of the categoricity result is then the claim that every reinterpretation of "0" and "*Pred*" that satisfies $BL(0, Pred)$ is isomorphic to every other, and in particular is isomorphic to the "real" $<0, Pred>$. The Fregean version is the claim (proven in G) that $\forall x \forall R(BL(x, R) \rightarrow ISO\ (<x, R>, <0, Pred>))$.

From a mathematical point of view, the difference between the two versions is insignificant: Frege's axioms are categorical in one sense iff they're categorical in the other, and for essentially the same reasons. For our purposes, the differences worth noting are as follows. First, Frege's version is not in any sense "metatheoretic." It's simply a derivation within *G* of a sentence expressing a universal claim about all objects and relations; it makes no mention of sentences or interpretations. Second, there are some in-principle differences between the modern and the Fregean approach to such results that don't show up in this example but are worth noting:

1. Because the Fregean method requires that the result to be proven be expressed as a conditional whose antecedent expresses the conditions laid down in the statements whose categoricity is in question (here, our "basic sentences"), the Fregean approach works only when such a conditional sentence can in fact be formed. If, for example, the basic sentences had been infinite in number, there would be no way of framing the categoricity result in *G* (or in any other finitary language). The modern method has no such restriction.
2. If one is working in a metatheory that's proof-theoretically stronger than the theory whose sentences are in question, it is possible for the modern result to be provable while the Fregean result is merely true but not provable.
3. If the metatheory has a richer domain than the object theory, the modern claim may be true while the Fregean claim is false. (Consider the case in which the object-language quantifies over a domain lacking the function whose existence is asserted in the isomorphism claim.)

That Frege's version of the theorem isn't metatheoretical is as one should expect, given Frege's interests and his view of the importance of the theorem. For Frege, the question of whether the basic laws he has proven will suffice for the rest of the arithmetic of the natural numbers would have been importantly bound up with the question of whether those laws sufficed to fix the overall structural features of that collection of numbers. The failure of the "isomorphism" result would have meant that some essential features of the collection of natural numbers had not yet been characterized by the basic laws proven to this point.[42] But this isn't, from Frege's point of view, naturally viewed as a point about other interpretations of the terms "0" and "*Pred*." It's rather a (very similar) point expressible by a sentence that explicitly quantifies over the positions held by those terms. This is just the difference between Hilbert's understanding of his consistency-proofs as involving reinterpretation and Frege's recasting of the point of those proofs as involving universally-quantified statements with fixed meanings.[43]

There is nothing to prevent Frege from stepping back and asking the modern isomorphism question. That he doesn't do so has to do, again, with (1) his historical

location, and (2) the fact that he does not think of the sentences themselves as the objects of interest. Taking sentences to be of interest only in virtue of their expression of determinate thoughts, questions about the models of a set of sentences are not natural ones to ask. Considerably more natural from this perspective are the roughly-equivalent Fregean questions of the truth of those universally-quantified thoughts whose antecedents are expressed by the conjunction of the sentences in question.

7.5 Conclusion

The metatheoretic issues with which Frege is explicitly concerned all have to do with the reliability of his formal system, and include (1) the reliability of individual rules and axioms, (2) the uniquely-referring nature of the sentences and subsentential pieces of language, and (3) the eliminability of defined terms. This is as we should expect. The formal system, from Frege's point of view, is a tool with a very specific purpose: to serve as a means for expressing proofs, a means that is guaranteed to present proofs in such a way that no step can "sneak in unnoticed"[44] in the chain of inferences from premises to conclusion. A proof expressed by a derivation in such a system will demonstrate, with as much certainty as is had by our conviction of the reliability of the system, that its conclusion follows logically from the premises explicitly laid out in that proof. A reliable such system of proof is just what Frege needs in order to demonstrate the logicist thesis. That Frege's only real metatheoretical concern is that of the reliability of his system is unsurprising: without the reliability of the formal system, the proof-theoretic project is bankrupt, and with it, the system meets the most important criterion for the use to which Frege wants to put it.

The other crucial criterion is that the system be sufficiently rich. That is, it is essential to Frege's logicist reduction that his formal system incorporate enough logical principles to carry out the necessary proofs. The concern with richness in this sense is again what underlies the move from *B* to *G*. If Frege had been right, first, that the whole of the arithmetic of the natural numbers was provable from his version of the Peano-Dedekind axioms and, second, that the system of *Grundgesetze* sufficed for the proof of those axioms, then the system *G* would have been sufficiently rich for the logicist project as applied to the theory of the natural numbers.

A stronger requirement is that of the comprehensiveness of a system, i.e. the condition that for each of its sentences ϕ and sets Σ of sentences, ϕ is derivable from Σ if the thought expressed by ϕ follows logically from the set of thoughts expressed by the members of Σ. From Frege's perspective, there is no systematic way to test whether a formal system is comprehensive, i.e. whether it incorporates all of the principles of valid logical inference applicable to the thoughts expressed

in its language. For there is no general testable condition met by all and only those pairs $<\Sigma, \phi>$ of a set of formulas and a formula, respectively, such that the first expresses a set of premises that logically entail the conclusion expressed by the second. No condition on such pairs plays for Frege the role that model-theoretic entailment plays for Tarski.[45] And, as we've seen, this is due not just to Frege's historical location, but also to philosophical principle: that a pair $<\Sigma, \phi>$ is (not) truth-preserving under every interpretation is not, for Frege, a guarantee that $\tau(\Sigma)$ logically entails (fails to entail) $\tau(\phi)$.

That Frege has no non-circular way of circumscribing the formulas that express truths of logic, or the pairs $<\Sigma, \phi>$ that express logically-valid arguments, means that he must adopt what's been called an "experimental" approach to the comprehensiveness of his system.[46] If we come across a principle of logical inference that isn't yet reflected in the formal system we're using, we simply add the relevant axiom or rule of inference to that system. Or, at least, we do so if we're aiming for comprehensiveness. It's worth recalling that the comprehensiveness of his formal system is by no means a requirement for the use to which Frege puts it. What's required for the logicist project is that the formal system reflect in its axioms and rules enough principles of logical inference to prove the fundamental truths of arithmetic. If there are further basic principles that aren't needed for arithmetic, their absence from the Fregean system won't be an impediment to the central use he makes of it.

To summarize: of the now-standard metatheoretic questions that Frege does not raise, some, most notably completeness, would have been of little interest from his point of view. Others, most notably soundness and consistency, would have been of considerable interest. Yet others, notably categoricity results, are close cousins to results that were of clear importance to Frege. None of these modern results would have been in any sense incoherent, and none would have been flawed in the way that, from Frege's point of view, Hilbert-style consistency-proofs for theories are flawed.

Notes

1. *Begriffsschrift* (hereafter *Bff*) Preface p vii/7.
2. *Grundgesetze* (hereafter *Gg*) 1 Introduction p vii/3–4. This passage is followed immediately by the sadly prophetic:

> A dispute can arise, so far as I can see, only with regard to my Basic Law concerning courses-of-values (V), which logicians perhaps have not yet expressly enunciated, and yet is what people have in mind, for example, when they speak of the extensions of concepts. I hold that it is a law of pure logic. In any event the place is pointed out where the decision must be made.

3. See *Bff* §24.

4. *Bff* §§5–6, emphasis added.
5. *Bff* §14, emphasis added.
6. *Bff* §14. Note the typographical error in the antepenultimate formula cited as it appears in van Heijenoort [1967b].
7. *Gg* 1 Introduction p xvii/15.
8. To say that an individual derivation-rule is *reliable* is to say that it sanctions the inference of a sentence ϕ from a set Σ of sentences only if the thought expressed by ϕ follows logically from those expressed by the members of Σ; similarly, an axiom-sentence is *reliable* if the thought it expresses is a truth of logic.
9. *Gg* 1 Introduction p xii/9. See also *Gg* 1 §5 footnote: "Denotationless names must not occur in the Begriffsschrift."
10. This is the argument whose conclusion is, "Thus without contradicting our setting '$\grave{\epsilon}\Phi(\epsilon) = \grave{\epsilon}\Psi(\epsilon)$' equal to '$\forall x(\Phi(x) = \Psi(x))$' it is always possible to stipulate that an arbitrary course-of-values is to be the True and another the False" (*Gg* 1 §10).
11. This is a slightly misleading way of putting it, since the term-forming operator of Frege's system refers to a first-level function; its ordinary use will be to operate on the course of values of a concept, yielding the single object falling under that concept (if there is one) and that course of values itself otherwise. (As applied to an object other than a course of values, it will return that object as value.)
12. This argument, as has been often remarked (including by Frege himself), is shown by Russell's paradox to be flawed, for the paradox demonstrates the existence of course-of-value names that, though well-formed by Frege's lights, do not receive a (unique) denotation via his stipulations. For Frege's reaction to this effect, see the letter to Russell June 22, 1902. For an enlightening discussion of these sections of *Grundgesetze*, see Heck [1997]. See also Resnik [1986], Parsons [1965].
13. "Namely, by our stipulations it is determined under what conditions the name denotes the True. The sense of this name—the *thought*—is the thought that these conditions are fulfilled" (*Gg* 1 §32).
14. *Gg* 1 §12, emphasis added. As Frege notes here, this is just one case of such interchangeability, each case of which ought in principle to be justified separately, but: "So as not to become tied up in excessive complexity, I here wish to assume this interchangeability generally granted, and to make use of it in future without further explicit mention." Ibid.
15. *Gg* 1 §14.
16. *Gg* 1 §15. I have suppressed Frege's indexes "(α" and "(δ", respectively, on the first two propositions quoted. The rule is called the "second" method of inference despite the previous introduction of both the interchangeability of subcomponents and modus ponens because Frege does not honor the first of these with the title "method of inference."
17. *Gg* 1 §18.
18. Dreben and van Heijenoort [1986] p 44.
19. Goldfarb [1979] p 353.
20. Ricketts [1998] p 174.
21. For a nice discussion of this issue, see Sullivan [2005].
22. Ricketts [1998] p 172.
23. Ibid., p 177.
24. As Ricketts puts it (ibid.):

> A formalized proof of the soundness of Frege's codification of logic would thus presuppose the truth of the axioms and the soundness of the inference rules of that codification. Given Frege's view of justification as explanation within unified science, such a proof would be scientifically pointless.

25. That is, the quick argument just outlined that gets one from exclusivism about *S* to the circularity of demonstrations of *S*'s reliability is not available in the case of demonstrations of *S*'s comprehensiveness, etc. This does not mean that there is no argument from this premise to the conclusion in question: one might well wonder whether the exclusivist

position makes all metatheoretic claims or demonstrations problematic on the basis of difficulties about self-reference, since from the exclusivist position it follows that arguments about *S* will be carried out in *S*. We turn to this question next.

26. [1882a] pp 97–98/90–91.
27. [1882b] pp 113–114/89.
28. Ibid.
29. *Gg* 1 Introduction p viii/4–5.
30. *Gg* 1 §0.
31. Ricketts and Weiner argue that the passages discussed in Section 7.1 are not in fact metatheoretical but are rather what Frege calls "elucidations," non-argumentative discussions intended to bring his audience to an understanding of the signs in question. See Ricketts [1986b] especially §2; Weiner [1990] Ch. 6. But this "elucidation" interpretation of the passages in question is not an easy one to maintain. First of all, there would seem to be no need at the relevant points in *Begriffsschrift* and *Grundgesetze* to try to bring the reader to grasp the meanings of the signs in question: all that is required for grasping these meanings has been already explicitly laid down in the introduction of those signs. Second, "elucidations" are, for Frege, what one gives when definitions are impossible, i.e. when the concepts one is trying to convey are too simple to admit of further analysis into previously-grasped concepts. Thus for example the distinction between function and object is one that, for Frege, can only be "hinted at" or gotten across via elucidation. See e.g. [1892c] para. 4. But there is no need for elucidation in the sections of *Begriffsschrift* and *Grundgesetze* in question: because one can say in German "the case in which *a* is affirmed and *b* is denied is denied," one can give explicitly, and without any need for elucidation, the meaning of "$(a \rightarrow b)$"; similarly for the additional primitive signs. For an argument to the effect that the passages in question are indeed justificatory, see Stanley [1996] and Heck [1997].
32. Ricketts [1985] p 11.
33. Ibid.
34. Ibid.
35. These linking principles can of course be *described* via straightforward statements, e.g. "the consequent of a conditional is logically implied by that conditional together with its antecedent." But to adopt such a statement as a premise is not thereby to introduce links where none existed already.
36. Ricketts [1986b] holds that a further reason to deny metatheoretic reasoning to Frege is that such reasoning requires us to make ineliminable use of the predicate "true," e.g. when affirming the view that the rules of inference are truth-preserving, and that Frege rejects such use of that predicate. But as Jason Stanley [1996] argues, the texts that Ricketts takes to support Frege's rejection of this employment of the truth-predicate simply do not seem to do so; Frege takes truth to be indefinable, but not for that reason unfit as the reference of a predicate. Further, as Ricketts [ibid] notes, Frege's own discussions of truth and of denotation are at odds with this account of Frege's conception of the truth-predicate. In short, it is very difficult to find in Frege's texts any clear reason for taking him to reject the meaningfulness of those metatheoretic claims whose expression employs the predicate "true."
37. For a clear example of a post-Hilbertian reading of the significance of completeness that's at odds with a Fregean understanding of the significance of that result, see Quine [1970].
38. That is, strictly speaking, soundness is a necessary condition for reliability if the system meets these conditions: (1) to each structure there is a corresponding acceptable reading; (2) *truth-on* is defined in such a way that a sentence is true on a structure iff it is true under the corresponding reading; and (3) derivability within the system is insensitive to the contents of non-logical terms.
39. It is worth remarking that the fact that Frege is irremediably wrong here, as demonstrated by Gödel's first incompleteness theorem, shows us that Frege's proof-theoretic method of demonstrating logicism is fundamentally flawed; it doesn't show us that the thesis itself is false (unless one is antecedently wedded to the idea, as Frege arguably was, that the truths of logic are recursively enumerable).

40. See, e.g., Parsons [1987]; also Heck [1996].
41. Heck [1993] in Demopoulos [1995] pp 283–284. The "members of the Q-series beginning with a" are just those things that bear the weak ancestral of $Q\xi\eta$ to a; "*Endlos*" is Frege's term for the first infinite cardinal.
42. This is not to say that the characterization of the order-type of the natural number series is all there is, from the Fregean point of view, to the successful characterization of those numbers; as we have seen, there is more to it than this. See Chapter 4 in this volume. It's a necessary but by no means sufficient condition.
43. It is perhaps worth pointing out here that the difference matters in the case of consistency-proofs in a way that it doesn't matter here, in the isomorphism case. As already discussed, the Fregean isomorphism-thought regarding *BL*(0, Pred) is true iff the modern isomorphism claim is true as well. In both cases, the question is essentially that of the strength of the defining conditions given by that collection of sentences, and comes down to the truth of a universally-quantified claim formed from those sentences. In the case of Hilbert's consistency-proofs, the issue is that of an existentially-quantified claim formed from the relevant sentences, and the implications thereof: where Hilbert sees a model, Frege sees a true existentially-quantified claim (so far, no significant difference); Hilbert then infers from this a consistency-result, which as we have seen simply does not follow, if "consistency" is understood in Frege's way. See Chapter 5 in this volume.
44. *Bff* Preface p 5.
45. This is a slightly-misleading way to put it. It's not the case that, for Tarski, a complete system encodes all principles of logical inference; note that e.g. a system of propositional logic can be complete in Tarski's sense. The point is simply that model-theoretic entailment provides from the Tarskian perspective something that it doesn't from Frege's: namely, a mark of logical entailment insofar as that can be cashed out for the language in question.
46. See van Heijenoort [1967a]; also Ricketts [1998].

8

Conclusion

Many of Frege's best ideas have survived the conceptual and technical developments of the last 100 years. His idea that we can so structure the expression of content that good steps of logical inference are encoded in syntactic transformation rules has turned out to be more significant than he could have imagined. He was by no means unique in pursuing the syntactic expression of logically-important properties of and relations between contents, but his combination of this general idea with his innovative analysis of content in terms of quantification, together with his beautifully-designed notation for the expression of content so analyzed, made his work a watershed in the development of logic. So too, Frege's analysis of the fundamental notions underlying a theory of, as he puts it, "following in a series" and of a theory of cardinality have survived to become entirely commonplace.

Some of the central Fregean ideas that haven't survived have perished for good reason. Frege's three-part combination of (1) logicism, (2) realism about arithmetical truth, and (3) the view that the principles of logic form a surveyable collection, together form a view that despite its initial plausibility we now know via the first incompleteness theorem to be clearly false. Whether Frege's logicist thesis itself—i.e. the thesis that the truths of arithmetic are each provable from purely-logical premises via purely-logical steps of inference—is itself plausible is a further question, and one not so easily settled. But one thing we know, again via incompleteness, is that Frege's means of demonstrating the thesis with a proof-theoretic reduction of the axioms of arithmetic can't work: there is no such manageable collection of axioms for arithmetic, again assuming roughly Fregean realism about arithmetic.[1] If the truths of arithmetic form a collection that Frege would have recognized, then logicism in his sense is plausible only if the truths of logic form a much less-manageable collection than he seems to have supposed; the collection must in particular not be recursively enumerable.

Frege's view that there are extensions—i.e. objects identifiable via the equivalence between sentences of the form "$\forall x(Fx \equiv Gx)$" and "the extension of F = the extension of G"—is decisively falsified by Russell's paradox. Also falsified along with it, arguably, is the very liberal kind of "re-carving" of content that was to have been

exemplified by this equivalence, the re-carving that gets us not just from co-extensiveness claims to extension-identity claims but also from claims about the parallelness of lines to claims of the identity of their directions and in general from claims of the form $\alpha R\beta$ to purported equivalents of the form $f(\alpha) = f(\beta)$. Russell's paradox doesn't immediately falsify the more innocent of these, but the failure of even one such equivalence undermines the idea that such object-revealing re-carvings are simply innocent reconceptualizations of a given content. A particularly interesting case of such a purportedly ontology-revealing equivalence is one that Frege seems to have appealed to everywhere, the presumed equivalence between claims of the form "sentence α is content-equivalent with sentence β," for various ways of cashing out the notion of content-equivalence, and their counterparts of the form "the thought expressed by α = the thought expressed by β." One question to raise about the general loss of innocence of the Fregean equivalence-claims, and the consequent failure of the general principle undergirding the treatment of the terms flanking the identity-sign as genuinely referring, is that of whether this failure undermines Fregean talk about thoughts. Of this, more below.

In addition to the Fregean principles that have clearly survived, and those whose rejection is well motivated by subsequent results, is the third and arguably most interesting camp, that of the principles that haven't survived but whose loss is due more to historical accident than to good reason. Here, centrally, is the Fregean idea that logical entailment is generally due not just to those features of contents that are explicitly represented in the syntactic structure of the sentences used to express them but also to properties of and relations between those contents that may have no syntactic marking in the collection of sentences at issue. The clearest implication of this view is that from the Fregean perspective, two sets of sentences can have radically-different syntactic properties and hence be "logically" inequivalent, where this is measured either model-theoretically or in terms of formal derivability, while expressing exactly the same set of thoughts and hence being, from Frege's point of view, logically indistinguishable. Similarly, two sets of sentences can be indistinguishable except for choice of atomic terms—and hence have all of the same formal-deductive and model-theoretic properties—and yet express sets of thoughts that have, from the Fregean point of view, significantly different logical properties. The simplest cases of these kinds of divergence between the logical relations as Frege understands them and the logical relations as cashed out in terms of derivability or of truth-preservation across structures appear when the logical complexity of some of the contents in question is masked by the use of atomic terms to express them.

The general idea here is not uniquely Fregean; we are all inclined to agree that, given a sufficiently coarse-grained representation of a theory, the syntax of the sentences used to express it may miss important underlying features of the logic of that theory. This is, for example, what we all take it is bound to happen if we restrict ourselves to the language of propositional logic, or of monadic predicate logic, for

the expression of mathematical content. It's also what obviously happens if we use simple terms for the expression of such logically-complex properties as that of *continuous function* or *cardinal equivalence*. But the familiar modern way of thinking of these logically-relevant but syntactically-suppressed relations of content is that, in ordinary circumstances, they can be ignored. That is, the familiar supposition is that, again in ordinary circumstances, the important logical relations are reflected in the syntax of the sentences in question. There are two ways of thinking of the rationale behind this supposition. The first is that, in ordinary circumstances, the analysis is finished: that all logically-relevant relations of content are reflected in the surface syntax. The second is that, in ordinary circumstances, the logical properties and relations we're interested in have nothing to do with the contents (analyzed or unanalyzed as they may be) of individual terms; the subject-matter in question is the kind of structure, or structure-defining condition, laid out by the schematically-understood sets of sentences in question. On either assumption, the question of consistency is, or is reducible to, the question of syntactic consistency and is immediately demonstrated by the construction of a model. Similarly, questions of logical entailment, on either assumption, are or are reducible to questions of syntactic derivability and are demonstrable either by the construction of a derivation or, in special cases, by the demonstration of model-theoretic entailment.[2]

From the Fregean point of view, consistency and logical entailment apply not to sets of sentences but to the sets of contents or thoughts they express; and, importantly, from this point of view, it is never legitimate to assume that the analysis is, in the sense just discussed, finished. That is, it's never legitimate to assume that the logically-relevant properties of and relations between the contents in question are all reflected in the syntax of the sentences at hand. And from this point of view, of course, Hilbert-style consistency proofs don't show consistency, and relations of logical entailment diverge significantly from relations of formal derivability and of model-theoretic entailment.[3]

Part of the issue just discussed is terminological: "consistent," in many modern quarters, simply means "has a model," and as we've seen, there are good reasons to use the term this way. For Frege on the other hand, "consistent" means something quite different, something that applies not to sentences but to what they express, and importantly something that isn't generally implied by the existence of a model of those sentences. But there is more than just a terminological quibble here: from the Fregean point of view, it is a mistake to think of logic as concerned just with syntactic structure, or just with those aspects of content that are represented thereby. A restriction of logic to these explicitly-represented features of content is from the Fregean point of view a restriction that misses much of what's important in logic. As we've seen, there can be for Frege logical entailment in cases in which the syntax of the sentences in question is insufficiently sensitive to reflect it, e.g. in cases in which, from the modern point of view, a *lack* of entailment is model-theoretically demonstrable. Here

one has to say that, despite the convenience of the modern approach, i.e. the neatness afforded by restricting attention to those aspects of content represented by syntactic form, nevertheless Frege was right: the syntactic consistency of a set, though an important feature in its own right, is no guarantee of the consistency, in the ordinary sense, of what's expressed by the members of that set. Similarly, the syntactic independence of a sentence from a set, though significant in its own right, is no guarantee that what's expressed by that sentence fails to follow logically from what's expressed by the members of that set.

In the same way, as we've seen, despite the independent interest of the relation of model-theoretic entailment, the Fregean has principled reasons for resisting the idea of this relation as an approximation of genuine logical entailment. Here the point is simply that model-theoretic entailment marks an important relation between defining conditions for classes of structures, but not in general, as we've seen, a robust or logically-significant relation between the truth-evaluable contents expressed by the sentences in question. Here the question of whether the Fregean is "right" is not a terribly clear one. The Fregean is certainly right to resist the idea that model-theoretic entailment coincides with any logically-significant relation between thoughts. On the other hand, the not-unreasonable idea that model-theoretic entailment is intended not to approximate such a relation but to mark off an important relation between sentences in terms of their structure-defining role makes the difference between the Fregean and the modern a matter of aiming at significantly different targets. There is of course no uniform intention behind the codification of the various model-theoretic relationships, including that of model-theoretic entailment, and hence no single correct answer regarding the extent to which the Fregean is in clear disagreement with the modern here. The most important point here is simply that the relations of entailment, consistency, inconsistency, etc. that the Fregean takes to be paradigmatically logical are not, for the most part, definable or analyzable in terms of model-theoretic properties or relations.

As we've noted earlier in this section, Russell's paradox arguably undermines the easy route that Frege seems to follow from talk about sentences to talk about contents or specifically about thoughts. Frege seems to shift effortlessly from the claim that each of a number of sentences is semantically similar in a certain way to the claim that there is some *thing* they have in common, namely, the thought that each expresses. Prior to Russell's paradox, this effortless shift may have seemed easily defensible: just as talk about directions is just talk about what parallel lines have in common, so too talk about thoughts is just talk about what certain sentences have in common. One of the lessons of the paradox is simply that such "re-carvings" are not, or at least not always, ontologically innocent: the identity-claim on the right-hand side cannot be simply a reconceptualization of the original equivalence-claim on the left, at least in problematic cases.

This difficulty with thoughts, together with general worries about a certain ontological profligacy involved in the supposition of such entities in the first place, gives rise to the question of whether the important distinctions between the Fregean and more-modern conceptions of logic turn on the Fregean commitment to these mysterious entities. Is it the case that in order to be roughly Fregean in one's conception of logic one must be committed to something like Fregean thoughts?

The answer here is a clear "no." It is easiest to present Frege's distinctive views about logical properties and relations as he does, and as we have done, in terms of thoughts. But it is important to see that none of the central views stressed in this volume, those that separate Fregean from later conceptions of logic, turn on the commitment to thoughts. Precisely the same points can be made by focusing on the idea of logical entailment and consistency as applying to (sets of) fully-interpreted sentences, independently of whether sentential interpretation is a matter of the expression of thoughts.

The distinctively Fregean view of consistency, put in terms of sentences, is just that the consistency of a collection of sentences turns on what those sentences say. From the Fregean point of view, it is significant to the issue of consistency what the individual non-logical terms mean, so that:

1. If two sets of fully-interpreted sentences differ just in the meanings had by some of their non-logical terms, then the consistency-questions as applied to these sets are distinct questions and may have distinct answers.
2. If the sentences in a set contain some uninterpreted, purely-schematic terms, then the issue of consistency can arise only for the higher-level relation defined by these sentences but not for the sentences themselves.
3. The consistency of a set of fully-interpreted sentences is not reducible to the satisfiability of the higher-level concept defined by those sentences when its terms are taken as place-holders.

In short, as long as one takes the *consistency* of a set of sentences to turn on what those sentences say, in a way that makes the meanings of non-logical terms relevant to the question of consistency, one is in a position to notice the large distance that Frege insists upon between the existence of a model of a set Σ and the consistency of that set.

Similarly, the idea that logical entailment applies just to fully-interpreted sentences, and that it does so in a way that's sensitive to the meanings of non-logical terms, provides the heart of the Fregean picture of entailment. Given such a picture, a reliable deductive system like that of Frege's *Begriffsschrift* is a good means of demonstrating relations of logical entailment, though of course not the converse: a demonstration of non-derivability is no guarantee of a failure of entailment, which is to say that it is no guarantee of independence. This

content-laden conception of logical entailment also underwrites just the gap to which the Fregean is committed between logical entailment and any relation definable in terms of structures. For again, on the view of entailment as turning in part on what's meant by the non-logical terms, sentences that are indistinguishable with respect to their relationship to structures for the language can play dramatically different roles with respect to logical entailment.

For Frege, the set of thoughts expressed by

$$(\Sigma_1) \quad \{\mathit{CardinalNumber}(\mathbf{1}), \forall y(\neg \mathit{Succ}(y, \mathbf{1}))\}$$

is logically inconsistent, which is to say that the thought expressed by its first member logically implies the negation of the thought expressed by its second. This is as compared with the set

$$(\Sigma_2) \quad \{\mathit{Horse}(\mathbf{Secretariat}), \forall y(\neg \mathit{Sire}(y, \mathbf{Secretariat}))\},$$

in which the second sentence, though false, is not logically contradicted by the first. The same result obtains if one takes the consistency- and entailment-questions to apply not to Fregean thoughts but to the fully-interpreted sentences themselves, as long as the meanings of the individual terms are reckoned relevant to these questions. That is, the Fregean will agree with the sentential theorist about the consistency- and entailment-questions as applied to Σ_1 and Σ_2 as long as there is agreement (1) that a correct analysis of zero, of successor, and of cardinal number will show that the members of Σ_1, so analyzed, logically imply a contradiction; and (2) that no such analysis will deliver such a result for Σ_2.

How, exactly, one is to draw the boundary between the truths about a subject-matter that are delivered by conceptual analysis (e.g., that cardinal numbers are essentially measures of cardinality) and those that aren't (e.g. that every horse has a sire) is a question Frege never directly addresses, or at least never with noteworthy success. In this, he is in good company: it is hard to say what, exactly, separates a good analysis from a failed attempt. The important point for Frege's own positive project is that there are in fact good analyses, many of which are straightforwardly recognizable as successful. And the important point for Frege's conception of logic is that what's revealed by a good analysis of a given subject-matter is relevant to the logical properties and relations borne by the truths about that subject-matter.

Notes

1. The role of the realism assumption here is just that with this assumption comes the conclusion that the theory of arithmetic, i.e. the set of true sentences of the language of arithmetic, forms a complete set.
2. The central such "special case" is that in which the formal system is complete.
3. To recap: given a reliable deductive system, the derivability of ϕ from Σ implies that $\tau(\Sigma)$ logically entails $\tau(\phi)$, but not vice-versa; see Section 4.6 in this volume. In special cases, e.g. in the case of ordinary first-order logic, the fact that Σ model-theoretically implies ϕ entails that $\tau(\Sigma)$ logically entails $\tau(\phi)$, but not vice-versa; see Sections 6.2 and 6.3 in this volume. More generally, outside the scope of special cases like that of a complete first-order system, the entailment holds in neither direction; see Sections 6.3 and 6.4 in this volume.

This page intentionally left blank

BIBLIOGRAPHY

Frege's Works Cited

[*Bff*] *Begriffsschrift, Eine Der Arithmetischen Nachgebildete Formelsprache Des Reinen Denkens*, Halle: Louis Nebert 1879.

English translation by Stefan Bauer-Mengelberg as: *Begriffsschrift, A Formula Language, Modeled Upon That of Arithmetic, for Pure Thought* in van Heijenoort [1967b] pp 5–82.

[*Gl*] *Die Grundlagen der Arithmetik, Eine logisch mathematische Untersuchung über den Begriff der Zahl*, Breslau: Wilhelm Koebner 1884.

English translation by J. L. Austin as: *The Foundations of Arithmetic, A logico-mathematical enquiry into the concept of number*, Oxford: Blackwell 1953.

[*Gg* 1] *Grundgesetze der Arithmetik Band I*, Jena: Hermann Pohle 1893.

Partial English translation by M. Furth as *The Basic Laws of Arithmetic*, Berkeley and Los Angeles: University of California Press 1967.

[*Gg* 2] *Grundgeseteze der Arithmetik Band II*, Jena: Hermann Pohle.

Some sections translated into English in Geach and Black [1980] pp 139–224.

[1879] "Anwendungen der Begriffsschrift," delivered before the Jenaische Gesellschaft für Medecin und Naturwissenschaft on January 24, 1879. Reprinted in [1998].

English translation as "Applications of the Begriffsschrift" in [1972] pp 204–208.

[1879-91] "Logik" in [1983] pp 1–8.

English translation as "Logic" in [1979] pp 1–8.

[1881] "Booles rechnende Logik und die Begriffsschrift," in [1983] pp 9–52.

English translation as "Boole's Logical Calculus and the Concept-script" in [1979] pp 9–46.

[1882a] "Über den Zweck der Begriffsschrift," delivered to the Jenaische Gesellschaft für Medicin und Naturwissenschaft on January 27, 1882. Reprinted in [1998] pp 97–106.

English translation as "On the Aim of the 'Conceptual Notation'" in [1972] pp 90–100.

[1882b] "Über den wissenschaftliche Berechtigung einer Begriffsschrift," *Zeitschrift für Philosophie und philosophische Kritik* 81 pp 48–56. Reprinted in [1998] pp 106–114.

English translation as "On the Scientific Justification of a Conceptual Notation" in [1972] pp 83–89.

[1885] "Über formale Theorien der Arithmetik," *Sitzungsberichte der Jenaischen Gesellschaft für Medecin und Naturwissenschaft, Jenaische Zeitschrift für Naturwissenschaft* 19, pp 94–104. Reprinted in [1990] pp 103–111.

English translation as "On Formal Theories of Arithmetic" in [1984] pp 112–121.

[1891] *Funktion und Begriff*, Jena: Hermann Pohle. Reprinted in [1990] pp 125–142.

English translation as "Function and Concept" in [1984] pp 137–156.

[1892a] "Über Sinn und Bedeutung," *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50. Reprinted in [1990] pp 143–162.

English translation by M. Black as "On Sense and Meaning" in [1984] pp 157–177.

[1892b] Review of Cantor's *Zum Lehre vom Trasfiniten, Zeitschrift für Philosophie und philosophische Kritik* 100 pp 269–272. Reprinted in [1990] pp 163–166.
English translation in [1984] pp 178–181.
[1892c] "Über Begriff und Gegenstand," *Vierteljahrsschrift für wissenschaftliche Philosophie* 16 192–205. Reprinted in [1990] pp 167–178.
English translation as "On Concept and Object" in [1984] pp 182–194.
[1892d] Draft of "Über Begriff und Gegenstand" in [1983] pp 96–127.
English translation in [1979] pp 87–117.
[1894] Review of Husserl's *Philosophie der Arithmetik I, Zeitschrift für Philosophie und philosophische Kritik* 103 pp 313–332. Reprinted in [1990] pp 179–192.
English translation in [1984] pp 195–209.
[1897a] "Logik" in [1983] pp 137–163.
English translation as "Logic" in [1979] pp 126–151.
[1897b] "Über die Begriffsschrift des Herrn Peano und meine eigene," *Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig* 48 pp 362–368. Reprinted in [1990] 220–233.
English translation as "On Mr. Peano's Conceptual Notation and My Own" in [1984] pp 234–248.
[1897c] "Begründung meiner strengeren Grundsätze des Definierens" [1983] pp 164–170.
English translation as "The Argument for my Stricter Canons of Definition" [1979] pp 152–156.
[1903] "Über die Grundlagen der Geometrie," *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12 pp 319–324, 368–375. Reprinted in [1990] pp 262–272.
English translation as "Foundations of Geometry: First Series" in [1984] pp 272–284.
[1906a] "Über die Grundlagen der Geometrie," *Jahresbericht der Deutschen Mathematiker-Vereinigung* 15 (1906) pp 293–309, 377–403, 423–430. Reprinted in [1990] pp 281–323.
English translation as "Foundations of Geometry: Second Series" in [1984] pp 293–340.
[1906b] "Kurze Übersicht meiner logischen Lehren" in [1983] pp 213–218.
English translation as "A Brief Survey of My Logical Doctrines" in [1979] pp 197–202.
[1906c] "Über Schoenflies: Die logischen Paradoxien der Mengenlehre" in [1983] pp 191–199.
English translation as "On Schoenflies . . ." in [1979] pp 176–183.
[1910] Frege's notes to Jourdain in Frege [1980] pp 179–206.
[1914] "Logik in der Mathematik," in [1983] pp 219–270.
English translation as "Logic in Mathematics" in [1979] pp 203–250.
[1918a] "Der Gedanke," *Beiträge zur Philosophie des deutschen Idealismus I* pp 58–77. Reprinted in [1990] pp 342–361.
English translation as "Thoughts" in [1984] pp 351–372.
[1918b] "Die Verneinung," *Beiträge zur Philosophie des deutschen Idealismus I* pp 143–157. Reprinted in [1990] pp 362–377.
English translation as "Negation" in [1984] pp 372–389.
[1923] "Gedankengefüge," *Beiträge zur Philosophie des deutschen Idealismus III* pp 36–51. Reprinted in [1990] pp 378–394.
English translation as "Compound Thoughts" in [1984] pp 390–406.
[1924/25] "Neuer Versuch der Grundlegung der Arithmetik" in [1983] 298–302.
English translation as "A New Attempt at a Foundation for Arithmetic" in [1979] pp 278–281.
[1972] *Conceptual Notation and Related Articles*, T. W. Bynum (ed) Oxford: Oxford University Press.
[1979] *Posthumous Writings*, Hermes, Kambartel, Kaulbach (eds) Chicago: University of Chicago Press. (Translation of most of Frege [1983].)
[1980] *Philosophical and Mathematical Correspondence*, Gabriel, Hermes, Kambartel, Thiel, Veraart (eds) Chicago: University of Chicago Press.
[1983] *Nachgelassene Schriften* (2nd revised edition), Hermes, Kambartel, Kaulbach (eds) Hamburg: Felix Meiner Verlag.
[1984] *Collected Papers on Mathematics, Logic, and Philosophy*, B. McGuinness (ed) Oxford and New York: Blackwell Press. (Translation of most of Frege [1990].)
[1990] *Kleine Schriften* (2nd edition), I. Angelelli (ed) Hildesheim: Georg Olms Verlag.
[1998] *Begriffsschrift und andere Aufsätze*, I. Angelelli (ed) Hildesheim: Georg Olms Verlag.

Other Works Cited

Beaney, M. and E. Reck [2005] *Gottlob Frege Critical Assessments of Leading Philosophers*, Oxford and New York: Routledge Press.

Benacerraf, P. [1981] "Frege: The Last Logicist," *Midwest Studies in Philosophy VI*, Peter French, et al. (eds), Minneapolis: University of Minnesota Press, 17–35. Reprinted in Demopoulos [1995] 41–67.

Blanchette, P. [1994] "Frege's Reduction," *History and Philosophy of Logic* 15, 85–103.

Blanchette, P. [1996] "Frege and Hilbert on Consistency," *Journal of Philosophy* 93, 317–336. Reprinted in Beaney and Reck [2005].

Blanchette, P. [2000] "Models and Modality," *Synthese* 124, 45–72.

Blanchette, P. [2007] "Frege on Consistency and Conceptual Analysis," *Philosophia Mathematica* 15(3), 321–346.

Blanchette, P. [forthcoming] "Frege on Shared Belief and Total Functions," *Journal of Philosophy*.

Boolos, G. [1985] "Reading the *Begriffsschrift*," *Mind* 94, 331–344. Reprinted in Demopoulos [1995] 163–181 and Boolos [1998].

Boolos, G. [1990] "The Standard of Equality of Numbers," *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge University Press, Cambridge MA, 261–277. Reprinted in Demopoulos [1995] and Boolos [1998].

Boolos, G. [1998] *Logic, Logic and Logic*, R. Jeffrey (ed), Cambridge, MA: Harvard University Press.

Boolos, G. and R. Heck [1998] "Die Grundlagen der Arithmetik, §§82–83," in Schirn [1998] 407–428.

Cantor, G. [1899] Letter to Dedekind July 28 1899, translated by S. Bauer-Mengelberg in J. van Heijenoort (ed) *From Frege to Gödel* Cambridge, MA: Harvard University Press 1967, pp 113–117.

Demopoulos, W. [1994] "Frege and the Rigorization of Analysis," *Journal of Philosophical Logic* 23, 225–246. Reprinted in Demopoulos [1995] 68–88.

Demopoulos, W. [1995] (ed) *Frege's Philosophy of Mathematics*, Cambridge, MA: Harvard University Press.

Dreben B. and J. van Heijenoort [1986] "Introductory Note to [Gödel] 1929, 1930, and 1930a" in *Gödel, Collected Works* vol. 1, Feferman et al. (eds), Oxford: Oxford University Press, pp 44–59.

Dummett, M. [1991a] *Frege Philosophy of Mathematics*, Cambridge, MA: Harvard University Press.

Dummett, M. [1991b] *Frege and Other Philosophers*, Oxford: Clarendon Press.

Dummett, M. [1991c] "Frege and the Paradox of Analysis" in Dummett [1991b] pp 17–52.

Etchemendy, J. [1990] *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press.

Evans, G. [1981] "Understanding Demonstratives" in Parrett and Bouveresse (eds) *Meaning and Understanding*, Berlin: DeGruyter, pp 280–303.

Evans, G. [1982] *The Varieties of Reference*, John McDowell (ed), Oxford: Clarendon Press.

Geach, P. and M. Black [1980] *Translations from the Philosophical Writings of Gottlob Frege*, Peter Geach and Max Black (eds), 3rd edition, Totowa, NJ: Rowman & Littlefield.

Gödel, K. [1930] "Die Vollständigkeit der Axiome des logischen Funktionenkalküls," *Monatshefte für Mathematik und Physik 37*, 349–360. Reprinted with translation in Gödel [1986] pp 102–123.

Gödel, K. [1986] *Collected Works, Vol. 1* edited by S. Feferman, J. Dawson, S. Kleene, G. Moore, R. Solovay, and J. van Heijenoort, Oxford University Press.

Goldfarb, W. [1979] "Logic in the Twenties: The Nature of the Quantifier," *Journal of Symbolic Logic* 44, 351–368.

Goldfarb, W. [1988] "Poincaré Against the Logicists" in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics, Minnesota Studies in the Philosophy of Science* vol. 11, Minneapolis: University of Minnesota Press pp 61–81.

Goldfarb, W. [2001] "Frege's Conception of Logic" in J. Floyd and S. Shieh (eds) *Future Pasts: The Analytic Tradition in Twentieth Century Philosophy*, Oxford: Oxford University Press, 25–41. Reprinted in Potter and Ricketts [2010] 63–85.

Heck, R. [1993] "The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*," *Journal of Symbolic Logic* 58, 579–601. Reprinted with minor revisions and a postscript in Demopoulos [1995] 257–294.

Heck, R. [1996] "The Consistency of Predicative Fragments of Frege's *Grundgesetze der Arithmetik*," *History and Philosophy of Logic* 17, 209–220.

Heck, R. [1997] "*Grundgesetze der Arithmetik* I §§ 29–32," *Notre Dame Journal of Formal Logic* 38(3), 437–474.

Hilbert, D. [1900] *Grundlegung der Geometrie*, Stuttgart: Teubner.

English translation of the 10th edition: *Foundations of Geometry*, L. Unger (trans), P. Bernays (ed). Open Court 1971.

Hodges, W. [2004] "The Importance and Neglect of Conceptual Analysis: Hilbert-Ackermann iii.3," in Hendricks et al. (eds) *First-Order Logic Revisited*, Berlin: Logos, pp 129–153.

Korselt, A. [1903] "Über die Grundlagen der Geometrie," *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12, 402–407.

Kremer, M. [2010] "Sense and Reference: The Origins and Development of the Distinction" in Potter and Ricketts [2010] 220–292.

Levine, J. [2002] "Analysis and Decomposition in Frege and Russell," *Phil Quarterly* 52(207), 195–216. Reprinted in Beaney and Reck vol. 4.

McGee, V. [1992] "Two Problems with Tarski's Theory of Consequence," *Proceedings of the Aristotelian Society* 92, 273–292.

Parsons, C. [1965] "Frege's Theory of Number," in M. Black (ed) *Philosophy in America*, Ithaca, NY: Cornell University Press, pp 180–203. Reprinted with postscript in Parsons, C. [1983]. Reprinted (with same postscript) in Demopoulos [1995].

Parsons, C. [1983] *Mathematics in Philosophy: Selected Essays*, Ithaca: Cornell University Press.

Parsons, T. [1987] "On the Consistency of the First-order Portion of Frege's Logical System," *Notre Dame Journal of Formal Logic* 28, 161–188. Reprinted in Demopoulos [1995] 422–431.

Poincaré, H. [1905] "Les Mathématiques et la Logique," *Revue de Metaphysique et de Morale* 13, 815–835.

Potter, M. and T. Ricketts [2010] (eds) *The Cambridge Companion to Frege*, Cambridge, UK: Cambridge University Press.

Quine, W. V. O [1970] *Philosophy of Logic*, Prentice-Hall.

Reck, E. and S. Awodey [2004] (eds) *Frege's Lectures on Logic Carnap's Student Notes 1910–1914*. This is an English translation of the German text of these notes edited and with an introduction by Gottfried Gabriel. Open Court Press, Illinois.

Resnik, M. [1986] "Frege's Proof of Referentiality" in *Frege Synthesized*, L. Haaparanta and J. Hintikka (eds), Boston: Dordrecht Reidel, pp 177–195.

Ricketts, T. [1985] "Frege, the Tractatus, and the Logocentric Predicament" *Noûs* 19(1), 3–15.

Ricketts, T. [1986a] "Objectivity and Objecthood: Frege's Metaphysics of Judgment," in *Frege Synthesized*, L. Haaparanta and J. Hintikka (eds), Reidel Publishing Company, pp 65–95.

Ricketts, T. [1986b] "Generality, Meaning, and Sense in Frege," *Pacific Philosophical Quarterly* 67, 172–195.

Ricketts, T. [1998] "Frege's 1906 Foray into Metalogic," *Philosophical Topics* 25(2), 169–188. Reprinted in Beaney and Reck vol. 2. Pagination is to this reprint.

Schirn, M. [1998] (ed) *The Philosophy of Mathematics Today*, Oxford: Clarendon Press.

Shapiro, S. [1991] *Foundations without Foundationalism*, Oxford: Oxford University Press.

Stanley, J. [1996] "Truth and Metatheory in Frege," *Pacific Philosophical Quarterly* 77, 45–70. Reprinted in Beaney and Reck vol. 2.

Sullivan, P. [2005] "Metaperspectives and Internalism in Frege," in Beaney and Reck vol. 2, pp 85–105.

Tappenden, J. [2000] "Frege on Axioms, Indirect Proof, and Independence Arguments in Geometry: Did Frege Reject Independence Arguments?" *Notre Dame Journal of Formal Logic* 41, 271–315.

Tappenden, J. [1997] "Metatheory and Mathematical Practice in Frege," *Philosophical Topics* 25, 213–264.

van Heijenoort, J. [1967a] "Logic as Calculus and Logic as Language," *Synthese* 17, 324–330.

van Heijenoort, J. [1967b] (ed) *From Frege to Gödel A Source Book in Mathematical Logic, 1879–1931*, Cambridge, MA: Harvard University Press.

Weiner, J. [1984] "The Philosopher behind the Last Logicist," in *Frege Tradition and Influence*, C. Wright (ed), Blackwell.

Weiner, J. [1990] *Frege in Perspective*, Ithaca, NY: Cornell University Press.
Weiner, J. [2005] "Semantic Descent," *Mind* 114, 321–354.
Weiner, J. [2007] "What's in a Numeral? Frege's Answer," *Mind* 116, 677–716.
Wilson, M. [1992] "Frege: The Royal Road from Geometry," *Nous*. Reprinted in Demopoulos (ed) [1995].
Wright, C. [1983] *Frege's Conception of Numbers as Objects*, Aberdeen University Press.

This page intentionally left blank

INDEX

Arbitrariness (of reference-assignments) 59, 83–89, 91–92, 98, 101–102
Arithmetic, ordinary content of 10–11, 23, 50, 57–70, 75
Axioms
  of *Begriffsschrift* and *Grundgesetze* 49, 152–153, 155–156
  of arithmetic 96, 99, 167–168
  of geometry 112–125

Basic Law V 94, 102–103, 107, 171
Begriffsschrift 7–8, 12–17, 29–31, 150–154
Benacerraf, P. 84, 105–106
"Boole's Logical Calculus and the Concept-Script" 18–19
Boolos, G. 26
Boolos, G. and Heck, R. 27, 107

Cantor, G. 148
Cardinal number 20, 23, 92–101
Caesar Problem 70–72
Carroll, L. 164
Categoricity 167–171
Cognitive criterion 33–35, 80–82
Completeness of a formal system 131, 141–142, 145, 149, 157, 165–167, 171
Comprehensiveness 159, 167, 170–171
Conceptual analysis 4–11, 23–25, 50, 77–105, 128–131, 180
  and consistency 108–131
  in *Begriffsschrift* 8, 13–17
  In *Grundlagen* 9–10, 19–22, 77–101
  in *Grundgesetze* 10, 22–23, 94–95
  in 1881 18–19
Consistency
  of a theory or a set of thoughts 108–131, 135–137, 144–147, 176–180
  of a formal system 167, 171
  syntactic 111, 120–123, 136–7, 144–147, 177–178
Content-carving 39–45, 89–95
Context principle 89
Course-of-value 23, 42, 58–9, 83, 88, 154

Definitions 12–14, 19, 21–22, 62–75, 112, 115–116
  as analytic 19
  as stipulative 12–13
  conditional 69–70, 73–74
  Hilbert's, in geometry 115–123
  piecemeal 62–68
Demopoulos, W. 26
Directions 39–45, 89–92
Derivation (distinguished from *proof*) 12, 22
  in *Begriffsschrift* 13, 15
Dreben, B. and J. van Heijenoort 157, 163
Dummett, M. 42–45, 53, 87–89, 106

Equinumerosity 20, 78, 94, 102
Etchemendy, J. 148
Evans, G. 52
Extension (of a concept) 20–22, 83, 91, 98, 102–103, 175

Finite cardinal numbers 20–22, 96, 100
Function 37
  total definition of 49, 55–75

Goldfarb, W. 26, 76, 157
Grundlagen 9–10, 19–22, 70–72, 83, 89–101
Grundgesetze 10, 17, 22–23, 31, 45–50, 58–62, 65–68, 72–75, 83, 96, 150, 154–156

Heck, R. 88, 106, 148, 168, 172–174
Hilbert, D. 109–131
Hodges, W. 131, 133
Hume's Principle 20–21, 88, 92–97, 99

Incompleteness
  of arithmetic 173, 175
  of systems of logic 142, 145, 165–166
Independence 108–131, 144–146

Jourdain, notes to 128, 133

Kant, I. 16, 26–27, 99
Korselt, A. 118
Kremer, M. 51

Liebmann, letter to 109–110, 118–119
Linguistic completeness 60–62, 65–70
Logical entailment 4–5, 108–109, 137–147
Logical laws 8, 13, 22, 42, 74–75, 127–128

McGee, V. 148
Model
  as a kind of reduction 84–87
  in consistency proofs 135–136
Model-theoretic entailment 137–147, 165, 174, 178

Non-referring terms 38–39, 49, 58–70, 154
(The) Number that Belongs to the Concept F 20–21, 70–71, 93–99

One 20, 95
"One is the successor of zero" 24, 80, 100
Ordering in a sequence 8, 12–17

Paradox of analysis 78
Parallel lines and directions 39–45, 89–92
Parallels axiom 111, 133–134
Parsons, C. 172
Parsons, T. 174
Peano, G. 56, 62, 69, 103–104
Peano-Dedekind Axioms 96, 99, 167–168
Poincaré, H. 26
Predication 37
Proof (distinguished from *derivation*) 12, 22

Quantifiers, range of 60–68, 72–75

Reading of a language 138–140
Reduction
  of arithmetic to logic 3, 7–11, 23, 102–105
  of theories in general 84–87
Reference
  indeterminacy of 58–59, 101–102
  of predicates 37, 55–75
  of sentences 37–39, 46, 55–60
  preservation of 82–87, 91–92, 98, 101–102
Reliability 139, 141, 150–156, 165–167, 170, 172
Resnik, M. 172
Ricketts, T. 133, 157–165, 173
Russell's paradox 102–103, 150, 172, 175, 178

Sense-Reference distinction 10, 22, 31–39
Sentences
  and structure 45–46
  of natural language 45–46, 55–75
  of ordinary arithmetic 9–11, 50, 56–75
  of *Begriffsschrift* and *Grundgesetze* 12, 45–50
  partially interpreted 46, 111–125, 144–145
Shapiro, S. 148
Sorites paradox 61, 69
Soundness 136, 165–167, 171
Stanley, J. 173
Stipulation 59
Successor 20, 95–96, 99–100, 167–168
Sullivan, P. 172

Tapenden, J. 133–134
Tarski, A. 171
Thoughts
  as bearers of logical relations 4–5, 28–29
  as content of ordinary arithmetic 28–29, 55–75
  as essential to Frege's conception of logic 179–180
  as expressed by sentences 28–36, 45–46, 60
  individuation of 31–36, 79–82
  role in geometry 112–125
  role in theories 4–5, 28–29, 31–32, 50
  structure of 39–46
Truth on a structure 140
Truth-value 37–39, 46, 55–60

van Heijenoort, J. 25, 174

Weiner, J. 25, 57–58, 76, 84, 106, 173
Wilson, M. 106

Zero 18, 20, 83, 95–96, 99–100, 107
ZFC2 143