THE KINDS OF MATHEMATICAL OBJECTS

BEAU MADISON MOUNT

Magdalen College

21 February 2017

D.Phil Thesis, Faculty of Philosophy

67,770 Words

ABSTRACT

The Kinds of Mathematical Objects is an exploration of the taxonomy of the mathematical realm and the metaphysics of mathematical objects. I defend antireductionism about cardinals and ordinals: the view that no cardinal number and no ordinal number is a set. Instead, I suggest, cardinals and ordinals are *sui generis* abstract objects, essentially linked to specific abstraction functors (higher-order functions corresponding to operators in abstraction principles). Sets, in contrast, are not essentially values of abstraction functors: the best explanation of the nature of sethood is given by a variation on the standard iterative account. I further defend the theses that no cardinal number is an ordinal number and that the natural numbers are, as Frege maintained, all and only the finite cardinal numbers. My case for these conclusions relies not on the well-known antireductionist argument developed by Paul Benacerraf, but on considerations about ontological dependence. I argue that, given generally accepted principles about the dependence of a set on its elements, ordinal and cardinal numbers have dependence profiles that are not compatible with any version of set-theoretic ontological reductionism. In addition, a formal framework for set theory with *sui generis* abstract objects is developed on a type-theoretical basis. I give a philosophical defence of the choice of type theory and discuss various questions relating to the nature of its models.

TABLE OF CONTENTS

- Abstract 2
- Acknowledgements 5
- Notice 9
- Introduction 10
  - 0.1 The Kinds of Mathematical Objects: A General Outline 10
  - 0.2 Some Assumptions and Limitations 14
- Chapter I: The Motley World of Mathematics 21
  - 1.1 The Challenge from Orthodoxy 21
  - 1.2 Mathematical Subfields, Assemblages, and Theories, and Kinds of Reduction 34
  - 1.3 Benacerraf's Argument 53
- Chapter II: Typed Frameworks 62
  - 2.1 Higher-Order Quantification and Faithful Model Classes 64
  - 2.2 Type Structures and Typed Languages 77
  - 2.3 Non-Extensional Models 91
  - 2.4 Tenability: Some Distinctions 104
- Chapter III: Three Paths of Abstraction 107
  - 3.1 Abductivist Abstractionist Realism: Characterizing the Position 110
  - 3.2 Neologicism and Content Recarving 116
  - 3.3 Trivialist Abstractionism 125
- Chapter IV: Dependence Arguments 132

- 4.1 Dependence, Ordinals, and Cardinals: The Main Argument 132
- 4.2 Ordinals and Cardinals by Abstraction 151
- 4.3 Integrating *Sui Generis* Cardinals and Ordinals with Set Theory 158
- 4.4 Ordinals, Cardinals, and Natural Numbers 173
- 4.5 Against *Ante Rem* Structuralism 184

• Chapter V: Dependence, Essence, and Grounding 192

- 5.1 Essence 193
- 5.2 Grounding 200
- 5.3 Schwartzkopff's Argument and Donaldson's Objection 206

• Chapter VI: Higher-Order Abstraction Principles 210

- 6.1 Basic Results: The Extensional Case 210
- 6.2 Philosophical Implications of the Extensional Case 214
- 6.3 M-Models and Higher-Order Intensionality 218

• Conclusion 223

• Bibliography 225

ACKNOWLEDGEMENTS

My greatest debt is to my supervisors, Volker Halbach and Timothy Williamson, for their profound intellectual and moral support during every stage of this project. The opportunity to work with two such philosopher-logicians over a period of five years at Oxford has been one of the greatest privileges of my life. Volker's mastery of the techniques of formal philosophy and sensitivity to its role and limits has influenced almost every page of this thesis; his work on truth is a model of what formal methods can accomplish. My interest in type theory began long ago when I read Tim's paper 'Everything', and has been sustained (along with many other philosophical projects) by our discussions ever since.

The Kinds of Mathematical Objects began its life as a B.Phil. thesis, *Numbers, Sets, and Higher-Order Frameworks*, supervised in its initial stages by Jeffrey Ketland. His critical engagement with and commitment to the project were invaluable.

I thank my examiners, Øystein Linnebo and Ian Rumfitt, for their extremely useful comments, and for their influence over the years in other contexts on my ideas about the nature of mathematical objects.

Many philosophers discussed the issues in the thesis with me and helped me come to grips with these topics. I particularly thank Andrew Bacon, Neil Barton, Corine Besson, Tim Button, Roy Cook, Fiona Doherty, Tom Donaldson, Cian Dorr, Kit Fine, Peter Fritz, Jeremy Goodman, Dan Isaacson, Paolo Mancosu, Carlo Nicolai, Lorenzo Rossi, James Studd, Alexander Paseau, Michael Potter, Agustín Rayo, Sam Roberts, Gillian Russell, Jeff Sanford Russell, Gabriel Uzquiano, and Dan Waxman. Special recognition is due to Jeremy, Peter, and Cian, who convinced me of the virtues of non-extensional type theory.

Others who were influential during the writing of this work include Simona Ai-

mar, Lucy Allais, Frank Arntzenius, Catrin Campbell-Moore, Sam Carter, Neil Dewar, Christina Dietz, Antony Eagle, Maegan Fairchild, John Hawthorne, Leon Horsten, Andrew Huddleston, Harvey Lederman, Annina Loets, Ofra Magidor, Toby Meadows, the late Josh Parsons, Karl Schafer, Thomas Schindler, Amia Srinivasan, Andrew Stephenson, Margot Strohminger, Natalia Waights Hickman, Juhani Yli-Vakkuri, and Andy Yu.

Two more Oxford people deserve special thanks: Johanna Schnurr, for her way of doing philosophy, precise but fearsomely imaginative, and for picnics and poetry; and Rachel Fraser, for sabre-sharp insights and a bevy of bespoke counterexamples, for Scotch and hours of sparkling discussion, for walks in Port Meadow with Wolsey. But most of all, I am grateful to them for two utterly irreplaceable friendships.

I came to know Corine Besson and Ana Mladenović Williamson in Dubrovnik years ago, and their friendship has been precious to me ever since. Joanne Diehl always has been and continues to be wonderful. I am grateful to Electra, excellent clamberer, *amie de la science et de la volupté*, for so much, and I shall always miss her profoundly.

The Princeton Philosophy Department was almost as important as the faculty of philosophy at Oxford in my intellectual evolution. I thank Dave Baker, Derek Baker, Karen Bennett, Lara Buchak, Adam Elga, the late Delia Graff Fara, Zena Hitz, Des Hogan, Steffi Lewis, Vera Kaufmann, Colin Klein, Tristram McPherson, Alexander Nehamas, Michael Smith, and Matt Strohl—and above all Dan Garber and Gideon Rosen, who had no reason to take a chance on a literature student who thought he could teach philosophy, but did so anyway.[1] From the Princeton years, I also wish to thank Brookes Brown, Katherine Brown, Lisa Cerami, Peter Fenves, Daniel Heller-Roazen, Brooke Holmes, Jess Nauright, May Mergenthaler, Juliet O'Brien, Sarah Pourciau, Andrea Ryan, P. Adams Sitney, and Amy Shuster.

But above all: Nick Stang, *bon vivant*, raconteur, and Kantian, who first convinced me of what has been confirmed many times since—that being an analytic philosopher does not require relaxing one's aesthetic commitments; Jack Woods, exemplary logi-

[1] Gideon deserves additional credit (or blame) for pointing out that, since one has few opportunities publicly to thank the people who matter in one's life, one should make the most of them (see Rosen 1992: ii). Hence the length of these acknowledgements.

cian, metaethicist, and general ne'er-do-well; and Matthew Moss, polymath, fellow defector, dining companion, and beloved friend.

I thank the following people from my time in Paris and Berlin: Jonathan Barnes, Jean-François Courtine, Thomas Kroedel, Tobias Rosefeldt, Marwan Rashed, Sandrine Sanos, Julie Singer, Barbara Vetter, Lisa Vogt, and Julia Zakkou. For seven years Kit Heintzman's love and wit have been a great comfort. Kit's habit of slashing away at academic pomposity like a charging hussar has also been a useful corrective to my very Oxonian vices.

I am indebted to too many people from my pre-philosophical life to list them all, but in particular I wish to thank Ian Baucom, Janelle Blankenship, Tom Boeker, Denis Donoghue, Joshua J. Friedman, Anne Garréta, Katherine Hagopian, Diana Hassenger, Fredric Jameson, John Henning, Julie Henning, Sean Lakey, Jason Long, James McCoy, Alice Kaplan, Stephen Katz, Joann Kleinneiur, Alison Lawrence, Cybelle McFadden, Michael McSwain, Kelly Mackinac Lux, David Mallard, Melissa Malouf, Erin Marsten, V.Y. Mudimbe, Sara Olack, Thomas Pfau, Wendy Piquemal, the late Reynolds Price, Eric Robinson, Andrea Schorr, Nancy Scott, Barbara Herrnstein Smith, Kenneth Surin, and Meg Watson.

Portions of this thesis were presented as talks at Pembroke Philosophy Club (Oxford University); the Seventh Cambridge Graduate Conference on the Philosophy of Mathematics and Logic; the Oxford–Bristol Graduate Workshop in Logic and Philosophy of Mathematics; the CRNAP Workshop Oxford/Paris/Princeton (École Normale Supérieure); the Mind, World, and Action Conference (IUC Dubrovnik); and the Fifteenth Congress of Logic, Methodology, and the Philosophy of Science (University of Helsinki). I am grateful to my audiences for their questions and comments. I also thank anonymous reviewers for *Thought: A Journal of Philosophy* and *Philosophia Mathematica*.

I am grateful to the Clarendon Fund, the Oxford Laboratory for the Philosophy and Psychology of Rationality and Decision, Magdalen College, St Catherine's College, and New College for financial support. I also particularly thank Philipp Koralus, who went to great lengths to find a place for me at the LPPRD and introduced me to many

fascinating aspects of the philosophy of reasoning in the process.

I thank my logic students at New College, St Catherine's College, and elsewhere. I hope I was of some use to them, but I am certain they were to me, since the best way to make oneself learn something is to be obliged to teach it.

I have been fortunate to have a wonderful extended family: in particular, I thank David Bezos, Ada Castilla, Jennifer Bezos, Michelle Bezos Neary, Rayssa Bezos, and above all Angel Bezos, whose unconditional support has been crucial. The absent ones are never absent from memory: the elder Angel and Georgina Bezos and my grandparents, Zoila Hernandez, Herman Mount, and Lacy Mount.

My debt to my parents, Kenneth and Sara Mount, goes far beyond what I can record here. Their tolerance for their only child's eccentricities has greatly exceeded what could reasonably be demanded of any parents. If (and I make no promises) the reader finds anything memorable in my choice of examples and allusions, my mother should be credited, who taught me that the traveller needs to observe the world closely; anything crisp and elegant in my prose, my father, who taught me that the writer needs to know when to leave things out. Sadly a thesis on the philosophy of mathematics is perhaps not the best place to display these lessons, but they were important ones nonetheless.

Finally: this thesis is dedicated to Catharine Diehl, who has been my closest friend and constant intellectual interlocutor for more than fifteen years. It was she who turned me into a philosopher, conversation by conversation—on midnight drives through the New Jersey sprawl, in Nick's flat during the heat wave in Paris, in Berlin cocktail bars, on hikes in Lofoten—and her comments, questions, and objections keep improving my work.

NOTICE

The Kinds of Mathematical Objects extends the work done in *Numbers, Sets, and Higher-Order Frameworks*, a B.Phil. thesis completed at Oxford University in June 2014. In particular, the content of pages 28–30, 58–60, 107–23, 133–37, 151–83, and 193–205 draws on portions of the earlier work, with revisions ranging in degree from very light to all but total. The remainder of the thesis covers topics not discussed in *Numbers, Sets, and Higher-Order Frameworks*.

Material on pages 87–96 will appear in 'Invariance without Extensionality', in *The Semantic Conception of Logic: Essays on Consequence, Invariance, and Meaning*, edited by Gil Sagi and Jack Woods (under contract with Cambridge University Press).

Material on pages 132–50 is under review at *Philosophia Mathematica*.

Material on pages 210–18 appeared in 'Higher-Order Abstraction Principles', *Thought: A Journal of Philosophy* **4** (2015): 228–36.

INTRODUCTION

0.1 The Kinds of Mathematical Objects: A General Outline

One of the key tasks of metaphysics is to give a systematic account of the sorts of things that exist. In this enterprise, it draws upon the deliverances of other sciences: we learn from physics, for example, that electrons form a kind, leptons a more general kind, elementary particles a still more general kind, and so on. It is a fundamental goal of metaphysics to integrate these results into a larger picture: the metaphysician considers the relationship not only among kinds such as *electron*, *lepton*, and *elementary particle*, but also among *physical object*, *spatiotemporally located entity*, *simple*, *composite object*, *substance*, *event* and so on. Even when the kinds in question are familiar and low in the hierarchy of genera, deep philosophical perplexities can arise: we are well-acquainted with persons and human animals, but it is far from clear whether the kind *person* is coextensive with the kind *human animal*.[1]

Where concrete objects are involved, taxonomic projects of this sort are a mainstay of contemporary metaphysics: many philosophers are engaged in figuring out where leptons and quarks, hawks and handsaws, persons and their bodies, events and actions, fit in the universal catalogue of things.[2] In contrast, relatively little attention has been paid to the taxonomy of mathematical objects. *The Kinds of Mathematical Objects* represents an attempt to fill part of this lacuna. In this work, I consider four of the most important sorts of mathematicalia: sets, cardinal numbers, ordinal numbers, and natural numbers. I defend *antireductionism* about cardinals and ordinals: the view that no cardinal number and no ordinal number is a set. Instead, I suggest, cardinals and ordinals are *sui generis* abstract objects, essentially linked to specific abstraction functors (higher-order functions corresponding to operators in abstraction principles). Sets, in contrast, are not essentially values of abstraction functors: the best explanation of the nature of sethood is given by a variation on the standard iterative account. I further

[1] See, for example, Olson (2007), Parfit (1984, 2012).

[2] For representative examples, see Grossman (1983), Hoffman and Rosenkrantz (1994), Baker (2000), Westerhoff (2005), Lowe (1989, 2006), Koslicki (2008).

defend the theses that no cardinal number is an ordinal number and that the natural numbers are, as Frege maintained, all and only the finite cardinal numbers.

Unlike the best-known challenge to reductionism, due to Paul Benacerraf (1965), my arguments will not be based on the existence of multiple, purportedly 'equally good' candidate ontological reductions. Instead, I shall focus on a distinctively metaphysical set of considerations, which turn on questions of *objectual dependence*. The notion of objectual dependence, which has been extensively studied in contemporary philosophy, will be discussed at length in §4.1. Here it suffices to note that the dependence of a set on its elements is commonly taken to be the paradigm of objectual dependence; I term a version of the principle the *strong basic datum* on dependence.[3] But, I contend, there are good reasons to believe that the dependence profiles of ordinal and cardinal numbers are not compatible with any version of set-theoretic ontological reductionism, given the strong basic datum. Thus, if these considerations are persuasive, no cardinal number is an ordinal and no cardinal or ordinal is a set.

This argument forms the central thread of Chapter IV. Before embarking on it, in Chapter I, I discuss the *prima facie* case for ontological reductionism, based on mathematical practice—a case that rests on the contention that set theorists' identification of the ordinals with von Neumann ordinals, and the cardinals with initial von Neumann ordinals, is persuasive evidence that the ordinals *genuinely are* von Neumann ordinals and the cardinals genuinely are initial von Neumann ordinals. I argue that, properly understood, mathematical practice does not impose this commitment: if anything, the evidence of mathematical practice outside set theory tells against reductionism. As for orthodox set theory itself, its success merely requires us to accept that the standard von Neumann reduction is *adequate as a proxy*—a constraint which a sophisticated antireductionist account has no difficulty meeting. After setting out a systematic account of mathematical theories and their relations, I discuss Benacerraf's multiple realizability argument and its limitations.

In Chapter II, I introduce some technical results and conventions that will be used

[3] See Fine (1994a, 1995b), Linnebo (2008).

later in this work. The official framework in which I formulate precise mathematical theories is a version of functional type theory, with higher-order quantifiers taken as primitive. I compare the system to alternative versions of type theory and defend it on philosophical grounds. Although the intended interpretation of this system is not model-theoretic, I discuss various kinds of models—both extensional and non-extensional—for this system and argue for transfer principles that relate these models to the intended interpretation (and thus allow us to draw conclusions about a relatively intractable higher-order realm on the basis of facts about more tractable models).

In Chapter III, I consider three accounts of the role played by abstraction principles in mathematics. The prevalent account is *neologicism*, developed initially by Crispin Wright and Bob Hale (Wright 1983, Hale 1987, Hale and Wright 2001b). In its technical approach, the neologicist programme is very similar to the version of non-reductionism I shall propose, but its philosophical motivations differ greatly. On the neologicist account, abstraction principles are quasi-definitional: they operate by 'recarving content'. I argue that there is no coherent notion of content that meets the requirements of neologicism. Nonetheless, I maintain that there is a fully justifiable form of *non-neologicist* abstractionism: *abductivist abstractionist realism* (AAR), on which abstraction principles have no special epistemological status, but express lawlike essence-capturing facts about mathematical objects. AAR is to be distinguished from *trivialist abstractionism* of the kind endorsed by Agustín Rayo (2013, 2016). Like the strong realist, the trivialist views the import of abstraction principles as metaphysical rather than epistemological, but he views them as higher-order identities (or something akin thereto)—a position which I claim is untenable.

In Chapter IV, I develop the dependence-based argument against the identification of cardinals and ordinals with sets and offer a positive antireductionist account. In particular, I give a novel characterization of the ordinals, using the abstraction principle RITA (Restricted Isomorphism-Type Abstraction). I argue that RITA is preferable to Roy Cook's similar ordinal abstraction principle SOAP (Cook 2003), because it generalizes naturally to a family of polyadic operators, RITA^n, for $n \in \mathbb{N}$. An elegant and power-

ful theory based on RITA^n can be developed, with sets alongside *sui generis* ordinals and cardinals; this theory, Modified Boolos Iterative Theory (MB), extends standard second-order Zermelo-Fraenkel set theory with choice and *Urelemente* (ZFCU^2). I argue that this picture of the set-theoretic universe is preferable to abstractionist accounts involving principles like George Boolos's NEW V. Then I provide an argument that no cardinal is an ordinal and that the natural numbers are exactly the finite cardinals; the theory that results from adding these principles to MB is Sets, Cardinals and Ordinals (SCO). Finally, I give arguments for preferring my *sui generis* antireductionist view to the *ante rem* structuralist account of the kind defended by Stewart Shapiro (1997).

Chapter V is an excursus, devoted to investigating whether a key premiss in the argument for antireductionism about cardinals can be defended using additional theoretical resources. The arguments in Chapter IV used only the notion of objectual dependence as a primitive; in Chapter V, I avail myself of other dependence-related concepts, such as *ground* and *essence.* Related arguments about cardinals and dependence have been given by Gideon Rosen and Steve Yablo (MS) and Robert Schwartzkopff (2011) and criticized by Thomas Donaldson (MS). I argue that my approach is more promising (and avoids an error made by Schwartzkopff) and is not subject to Donaldson's objection. Ultimately, however, I conclude that the two-place dependence relation among objects is better understood than the hyperintensional essence and grounding operators, and relatively little is gained by introducing them.

In Chapter VI, I consider how RITA^n can be extended into higher types and discuss the problem of higher-type abstraction principles more generally. Building on results from a paper by Roy Cook (2009), I show that one natural system of generalizations, D-RITA$^{n(i)}$, provided by the *deep higher-order abstraction principles* (DHAPs) corresponding to RITA^n, leads to inconsistency; the same applies to the DHAPs corresponding to other plausible base abstraction principles. In contrast, the model existence situation is better for *shallow higher-order principles* (SHAPs). As for Cook's original result, there is a complication: it need not transfer to non-extensional settings. I explore an assumption about the metaphysics of higher-order logic—which I term NONINFLATION—that would

lessen the danger.

My defence of these positions is tentative. My arguments for antireductionism rely primarily on abductive considerations: I contend that, when we attempt to integrate mathematics into a global theory that accounts for ontological dependence, modality, and other metaphysically relevant features of reality, we find reductionism to be less convincing than its competitors. But this is far from a knock-down argument, and weighing relevant criteria of theory choice is a difficult process. In particular, intuitions about dependence are not entitled to absolute deference. In the Conclusion, I discuss how we should go about trying to evaluate the overall force of the arguments I give for antireductionism.

0.2 Some Assumptions and Limitations

Needless to say, this work does not survey every issue in the philosophy of mathematics relevant to the project of mathematical taxonomy. I take several controversial claims for granted throughout the work. First of all, I do not engage with nominalist challenges to the existence of mathematical objects. In the classification of (informal) mathematical theories developed in §1.2, I allow for purely instrumentalist and fictionalist construals of individual theories, but I start from the presumptions that the fields of mathematics central to this study are to be given realist interpretations and that those interpretations will entail the existence of at least some mathematical objects. The main arguments for this position—based on the indispensability of mathematical objects to the best current theories offered by the natural sciences, from the claims of mathematics as a science autonomous unto itself, from our apparent perception-like access to mathematical objects, and from the seemingly Moorean status of the deliverances of the most basic parts of mathematics—have been exhaustively developed elsewhere in the literature, as have responses to the most important nominalist counterarguments.[4]

[4]For the first, the classic sources are Quine (1948, 1960b, 1981b) and Putnam (1971); for a comprehensive presentation, see Colyvan (2001). For the second, see Maddy (1997: 183–205). For the third, see Gödel (1964: 266–69) and Maddy (1990: 36–80). For the fourth, see Lewis (1991: 57–59) and Burgess

Secondly, I presuppose that classical logic is the correct logic for reasoning about both mathematics and the world in general (at least in the sense that all theorems of classical logic are true). Thus I shall not consider arguments about the use of intuitionistic logic in set theory (Dummett 1963, 1993c; Lear 1977; Tait 1998; Paseau 2001, 2003; Rumfitt 2015: 263–320). The intuitionist position is of great interest, but—unlike the supervaluationist variants of classical logic discussed below—it is extremely hard to motivate on the fundamentally realist picture of mathematics presumed here. As Ian Rumfitt notes:

> Dummett takes it to be obvious that mathematics does not describe a realm wholly independent of human thought. No one, he opines, could be so much of a realist as to hold that it did (Dummett [1993a]: 75). Many mathematicians and philosophers have in fact taken the position that Dummett rules out *ab initio* but, for present purposes, we may take the incorrectness of this extreme realist position as a premiss. (Rumfitt 2015: 265)[5]

This work presumes the 'extreme realist' position—the negation of the claim Dummett considers obvious. I do not claim that extreme realism is obvious—at least, it is far from obvious that it is obvious—but a defence of it will have to wait for another occasion. I also do not consider predicativist restrictions on higher-order logic—systems in which comprehension principles are restricted; in §2.1 I briefly explain why such restrictions are incompatible with a suitably strong account of higher-order quantification.[6]

(1983). The main nominalist arguments against realism about mathematicalia focus on the causal theory of reference, epistemic inaccessibility, the seeming queerness of abstract objects, and ontological parsimony. See Field (1989: 27–25) and Cheyne (2001) for a bevy of nominalist arguments based on the first three considerations, as well as Benacerraf (1973: 671–73) on the second and Kitcher (1983) on the third. On ontological parsimony, see Goodman and Quine (1947: 105). Realist rejoinders involve establishing counterexamples to overly stringent causal criteria, appealing to externalist accounts of knowledge unburdened by spatiotemporal access requirements, using comparisons with the posits of contemporary physics to disarm queerness considerations, and pointing to theoretical virtues that compete with parsimony—virtues at times neglected by philosophers with an excessive fondness for desert landscapes. For responses to claims about causal requirements, see Steiner (1975: 109–37) and Burgess and Rosen (1997: 28–49); for the comparison with physical objects, see Gödel (1964: 268) and Maddy (1990: 50–67); on relevant theoretical virtues, see Burgess and Rosen (1997: 205–43) and Maddy (1997: 206–232).

[5] Internal citation altered for uniformity with this work's bibliography.

[6] See Dummett (1994) for motivations and Linnebo (2016) for a discussion of abstraction principles in a predicativist setting. The governing picture is similar, although not identical, to predicativist restrictions on theories of second-order arithmetic discussed; see Feferman (2005) for an introduction to this subject.

Alongside a general acceptance of classical logic, I assume that absolutely unrestricted quantification is coherent. In general, in the absence of indications to the contrary, the first-order quantifier in both English and the formal systems I develop is to be read as ranging over absolutely all objects. Similar considerations apply at higher order, although here the details of the interpretation are much more delicate; I discuss these issues in §2.1. In other words, I endorse *generality absolutism*. The arguments between generality absolutists and *generality relativists*—deniers of the possibility of absolutely unrestricted quantification—are by now familiar.[7] To my mind, the most compelling consideration in favour of generality absolutism is the difficulty of coherently formulating its negation: as Lewis notes, the relativist who claims that *every* quantifier range is restricted, if he is to express what he wants, "violates his own stricture in the very act of proclaiming it" (1991: 68).[8] This is not absolutely decisive: that we cannot express a claim does not entail that it is false. If there were compelling reasons to reject generality absolutism, we could coherently retreat to what Tim Button (2010) terms "militant quietism", avoiding self-undermining formulations of generality relativism simply by refusing to assert any position on the issue.

But there are no such compelling reasons. The central argument for generality relativism, due to Dummett (1981: 475–76; 1991: 316–17) among others, is that the lesson to be drawn from Russell's Paradox is that the concepts associated with certain sortals ('set', 'ordinal', 'thing', and so on) are *indefinitely extensible*: where F is such a sortal, any attempt to quantify over all Fs both picks out a domain $\mathscr{D}$ of Fs and provides a recipe for specifying an F that does not fall within $\mathscr{D}$; adjoining this object to $\mathscr{D}$ generates a more capacious domain, $\mathscr{D}^+$, which in turn yields provides a recipe for specifying an F that does not fall within $\mathscr{D}^+$, and so on. But this picture of quantification assumes that, in order to quantify over some things, there must be *one thing*, some set or collection (the domain) within which they are all contained and over which the quantifiers can range—an assumption that Richard Cartwright (1994: 7) terms the 'All-in-One Princi-

[7] See Cartwright (1994), McGee (2000), Williamson (2003), and the essays in Rayo and Uzquiano (2006).

[8] This argument is developed at length in Williamson (2003: 427–35).

ple'. It is open to the generality absolutist to deny the All-in-One Principle and reject indefinite extensibility. This is the position taken here. Denying the All-in-One Principle means that reference to classes and other larger-than-set-sized collections can only be a *façon de parler*; in §2.1, I discuss how to regiment these idioms, in an internally coherent fashion, in higher-order terms.[9]

Thirdly—and more tentatively—I assume that the correct semantics for our most general language for serious theorizing about the world is bivalent.[10] Here my position is hedged: I allow that there may in principle be *multiply interpreted* mathematical theories, and I argue that the most plausible (although not the sole feasible) semantics for such a theory is supervaluationist (and thus violates bivalence). Although every theorem of classical logic is a theorem of supervaluationist logic, the supervaluationist consequence relation becomes deviant in the presence of a 'definitely' operator.[11] Since such an operator is usually not required for mathematics or the metaphysics of mathematical objects, however, we can think of ourselves as working in a fully classical fragment.

It is my contention that *in fact* we should not take our fully systematized theories of sets, cardinals, ordinals, and natural numbers to be multiply interpreted (whether or not reductionism is true): the costs of abandoning bivalence outweigh the benefits. But I admit that there are two cases where multiply interpreted accounts of the mathematical theories under consideration are sufficiently plausible that the matter requires more detailed treatment: pluralist treatments of set theory and supervaluationist structuralist treatments of cardinals, ordinals, and the natural numbers.

The general view of set theory taken here is *universist*: I presume that we have sufficient grasp of elementhood and sufficient quantificational resources to pin down

[9] See also Uzquiano (2003) on a related treatment of classes in plural logic.

[10] If bivalence is understood as the claim $\forall x(\mathsf{Sent}(x) \to Tx \vee T\dot{\neg}x)$, then exceptions may be required, since this principle fails in reasonable untyped truth theories. See Halbach (2014) for details. If truth is treated as a defined notion in terms of propositional quantification (Williamson 1998, Andjelković and Williamson 2000), then bivalence should plausibly be formulated as the weaker claim $\forall x(\mathsf{Sent}(x) \wedge \exists P\,\mathsf{Says}(x,P) \to Tx \vee T\dot{\neg}x)$, which can be implemented unproblematically even in theories allowing for self-application of the truth predicate.

[11] See Williamson (1994), Varzi (2007).

fully a unique intended interpretation both of the informal set theory we use when doing ordinary mathematics and the interpreted formal theories of sets, such as ZFC(U) or my SCO, that regiment that informal theory. In contrast, *pluralists* maintain that there are multiple equally good interpretations of set theory. Generality absolutism fits naturally with universism about set theory, although it does not entail it. (Later on, I frame my discussion of multiply interpreted theories, which can in principle be applied to set theory, using absolutely unrestricted (higher-order) quantification in the metalanguage, applying the notion of an *open-ended class of interpretations*, of which the generality relativist, who rejects absolutely unrestricted quantification, cannot avail herself.)

In §1.2, I discuss briefly the considerations which I take to favour universism over pluralism. Nonetheless, the antireductionist arguments presented in this work do not require universism as a premiss. In particular, the arugments in §4.1 that ordinals and cardinals are not sets could be carried out entirely within first-order ZFCU (supplemented with additional nonlogical predicates). Hence they use only resources acceptable to someone who endorses a multiply interpreted set theory (provided that her theory extends ZFCU): although a set-theoretic pluralist may reject the positive account of *sui generis* cardinals and ordinals I offer and the second-order system SCO that formalizes it, she cannot avoid confronting the antireductionist argument itself. This is particularly clear if a multiply-interpreted set theory is given a supervaluationist semantics: the argument will use only supertrue premisses.[12]

A similar situation arises for *multiple-reduction structuralism* about natural numbers, cardinals, or ordinals: dependence-based arguments telling against this view can be carried out without begging the question. In §1.2, I discuss this position in more detail, but the general idea is simple: a multiple-reduction structuralist might hold, for example, that although the natural numbers are pure sets, there is no one particu-

[12] On pluralism generally, see Hamkins and Linnebo (MS). In Linnebo (2013), a potentialist version of set-theoretic pluralism (see §1.2), is formalized using a modal operator that obeys S4.2 + CBF; a similar approach is taken in Studd (2013). Note that arguments with supertrue premisses are both globally and locally valid, so the precise notion of validity used in the supervaluationist semantics is not important; see Williamson (1994: 147–49) for discussion.

lar collection of pure sets determinately containing all and only the natural numbers. Instead, every ω-sequence of pure sets corresponds to an appropriate way of completing the incomplete determination imposed on the natural numbers by the combination of pure-set reductionism and the structural properties shared by all ω-sequences. If multiple-reduction structuralism is combined with a supervaluationist semantics, then (for example) '$2 < 3$' will come out supertrue, but '$2 \in 3$' will come out neither supertrue nor superfalse. The resulting combination provides one natural response to Benacerraf's antireductionist argument. Such a combination, however, remains vulnerable to antireductionist arguments of the sort developed here. Once again, every premiss in the arguments in §4.1 can be defended as supertrue even in the framework of a multiple-reduction account. Thus, although I maintain that there are independent reasons to reject multiple-reduction theories, I do not appeal to these reasons in my antireductionist arguments.

In addition, I make no attempt to canvass the entire variety of structuralist options. The central competitor to my view is Stewart Shapiro's *ante rem* structuralism (1997: 84–106); I argue against it in §4.4. I also give arguments in passing against common versions of eliminative structuralism. But I do not consider *modal* structuralisms of the type defended by Geoffrey Hellman (1989, 2001: 198–205).[13] My considered view is that the non-metaphysical modal operators used in Hellman's system are irremediably obscure, unless they are interpreted as effectively quantifying over interpretations of the quantifier in mathematical settings—in which case the view collapses into a multiple-interpretation account of the kind discussed in §1.2. Again, however, considerations of space preclude defending this claim here.

Finally, a word about notation and terminology. Specialized notation is introduced as it is used, but in the informal set theory used as a metatheory, the following conventions apply: The class of functions from A to B is denoted ${}^{B}A$. The set-theoretic difference of A and B is denoted $A - B$. We normally use $\mathrm{card}(A)$ or $|A|$ for the cardinality of A, reserving # and kindred operators for formal systems; context will determine

[13] For more on modal structuralism, see Parsons (1990: 318–21).

whether the true cardinal or the von Neumann cardinal is meant (although in mathematical, rather than philosophical, settings it will make little difference). $\mathscr{P}_{<\kappa}(A)$ is the class of subsets of A of cardinality $< \kappa$, and $\lambda^{<\kappa} = |\mathscr{P}_{<\kappa}(A)|$ for some arbitrary A such that $|A| = \lambda$. The capitalized terms 'Concept', 'Superconcept', 'Relation', and variants are used as Fregean shorthand: the details are discussed in §2.1. Their lower-case analogues have their usual meanings. I use the term 'functor' for a higher-order function; the semantic value of an abstraction operator is a functor. An abstract is the value of an abstraction operator; I use 'abstractum' as a general term for an object that is abstract in the broader sense (non-spatiotemporal, bereft of causal powers, and so on).

I: THE MOTLEY WORLD OF MATHEMATICS

1.1 The Challenge from Orthodoxy

The pressure in favour of set-theoretic ontological reductionism, at first glance, seems strong. After all, set-theoretic reductionism appears to be the answer to the question of mathematical taxonomy implicit in virtually every contemporary work on set theory: the ordinal numbers are identified with the von Neumann ordinals, the cardinals with initial von Neumann ordinals, and the natural numbers with finite von Neumann ordinals.[1] Collectively, we can term these theses *full von Neumann reductionism*.

It will be useful to develop a standard notation for these claims. We use U and V for the universal class and the class of pure sets, respectively.[2] We use On, Card, and $\mathbb{N}$, for the class of true ordinals, the class of true cardinals, and the class of true natural numbers, whatever they turn out to be, with the following notation for their members:

$$\mathsf{On} = \{0_{Ord}, 1_{Ord}, \ldots, \omega_{Ord}, \omega + 1_{Ord}, \ldots, \omega_{1Ord}, \ldots\};$$

$$\mathsf{Card} = \{0_{Card}, 1_{Card}, \ldots, \aleph_{0Card}, \aleph_{1Card}, \ldots\};$$

$$\mathbb{N} = \{0_{Nat}, 1_{Nat}, \ldots\}.$$

(Normally I write $\aleph_0$ and ω, rather than $\aleph_{0Card}$ and ω_{Ord}, since the standard notation provides sufficient disambiguation.)

When there is no danger of confusion, subscripts will be left off terms for elements of $\mathsf{On}, \mathsf{Card}$, and $\mathbb{N}$ (i.e., the *true* ordinals, cardinals, and natural numbers) and variables ranging over them.[3] We assume that each of $\mathsf{On}, \mathsf{Card}$, and $\mathbb{N}$ is equipped with a standard ordering, which we write $<_{Ord}$, $<_{Card}$, and $<_{Nat}$, respectively. Normally, however, we simply write $<$, relying on context for disambiguation; we introduce analogously standard notation for addition, multiplication, exponentiation, and other op-

[1] See, e.g., Kunen (1980, 2013), Drake (1974), and many others.

[2] As mentioned, all such informal class talk is ultimately to be regimented in a higher-order form. Here, however, I shall avoid undue punctiliousness. I also use set-builder notation for class terms and write $A = \{a, b, \ldots\}$, even when A is not countable, when no confusion will result.

[3] The one-place predicates 'Ord', 'Card', and 'Nat' will also be used in regimented contexts.

erations. We use $\alpha, \beta, \gamma, \ldots$, and decorated variants for variables ranging over true ordinals (or, occasionally, schematic letters for a specific ordinal); $\kappa, \lambda, \mu, \ldots$, and their variants play the same role for cardinals, as do $i, j, k, m, n, \ldots$, and their variants for natural numbers.

We define the *von Neumann ordinals* as $\mathsf{On}_{VN} = \{x \in V : \text{Trans } x \wedge \text{Connex } x\}$, using standard abbreviations: Trans x stands for $\forall y \forall z(y \in z \wedge z \in x \rightarrow y \in x)$ and Connex x stands for $\forall y \forall z(y \in x \wedge z \in x \rightarrow y \in z \vee y = z \vee z \in y)$.[4] We further introduce terms for the individual von Neumann ordinals as follows, with notation for constants and variables differing from that for true ordinals by subscript: $\mathrm{o}_{VN} = \emptyset; (\alpha + 1)_{VN} = \alpha_{VN} \cup \{\alpha_{VN}\}$; for limit δ, $\delta_{VN} = \bigcup_{\alpha < \delta} \delta_{VN}$; $\mathsf{On}_{VN} = \bigcup_{\alpha \in \mathsf{On}} \alpha_{VN}$.[5] We define $<_{VN}$ in the standard way, $\mathsf{Card}_{VN} = \{x \in \mathsf{On}_{VN} : \forall y \in \mathsf{On}_{VN}(y < x \rightarrow y \not\approx x)\}$ (where $\approx$ abbreviates a formula for equinumerosity, which can here be taken as primitive or defined, either in higher-order terms or using first-order set-theoretical resources, with no effect on the end result) and $\mathbb{N}_{VN} = \{x \in \mathsf{On}_{VN} : x < \omega_{VN}\}$, where $\omega_{VN} = \bigcap\{x \subseteq \mathsf{On}_{VN} : \emptyset \in x \wedge \forall y(y \in x \rightarrow y \cup \{y\} \in x)\}$. Thus, full von Neumann reductionism is the conjunction of the following claims: $\mathsf{On} = \mathsf{On}_{VN}$, $\mathsf{Card} = \mathsf{Card}_{VN}$, and $\mathbb{N} = \mathbb{N}_{VN}$. We also define, for comparison purposes, the *Zermelo natural numbers* as follows: $\mathbb{N}_Z = \bigcap\{x \subseteq \mathsf{On}_{VN} : \emptyset \in x \wedge \forall y(y \in x \rightarrow \{y\} \in x)\}$; notation for individual Zermelo natural numbers is introduced analogously to that for von Neumann ordinals: $\mathrm{o}_Z = \emptyset; (i + 1)_Z = \{i_Z\}; \mathbb{N}_Z = \bigcup_{i \in \mathbb{N}} i_Z$. *Zermelo reductionism* is the thesis that $\mathbb{N} = \mathbb{N}_Z$. More generally, *pure set ontological reductionism* about a mathematical kind K is the theory that $\|K\| \subseteq V$ (where $\|K\|$ is the class corresponding to K); *total pure set ontological reductionism* is the thesis that pure set ontological reductionism holds for every mathematical kind, or equivalently that the mathematical universe is identical to V.

With the von Neumann definitions in place, integers can be defined as appropriate

[4] The reduction is standardly attributed to von Neumann (1923), but it was independently discovered by Zermelo; see Ebbinghaus and Peckhaus (2007: 133).

[5] If in fact von Neumann reductionism about ordinals is false, then these clauses represent a noncircular definition of On_{VN} by transfinite recursion on the elements of On; if von Neumann reductionism is true, then the clauses cannot serve as a definition, but they nonetheless illuminate the structure of the sets in On_{VN}.

equivalence classes of (Wiener-Kuratowski) ordered pairs of natural numbers (under the equivalence relation $\langle a,b\rangle \sim \langle c,d\rangle \leftrightarrow a+d=b+c$, for example), rational numbers can be defined as appropriate equivalence classes of ordered pairs of integers (under the equivalence relation $\langle a,b\rangle \sim \langle c,d\rangle \leftrightarrow ad=cb$, for example), real numbers as Dedekind cuts, complex numbers as ordered pairs of real numbers, and so on.[6] And this is standardly how real numbers, complex numbers, and similar mathematicalia are introduced in top-down presentations of mathematics from the standpoint of set theory (*e.g.*, Enderton 1977).

Given this fact, it might be wondered whether any version of ontological antireductionism could be appropriately deferential to mathematical practice. As David Lewis (1991: 58) famously remarked, "Mathematics is an established, going concern. Philosophy is as shaky as can be. To reject mathematics for philosophical reasons would be absurd."[7] He made this observation in the context of a proposal to reject set theory, but does it not also apply to proposals to reject von Neumann reductionism?

The question is particularly important for the evaluation of arguments of the kind that will be advanced here. The chief argument of Chapter IV makes use of a notion of primitive ontological dependence. I claim that we can make sense of this notion and that we have reason to consider our intuitions about its extension broadly reliable, much as we have reason to consider our intuitions about essence facts or the contours of modal space broadly reliable. But this is nothing like the epistemic position we occupy with respect to the ZFC axioms themselves. The kind of metaphysical inquiry I shall engage in cannot without hyperbole be called as shaky as can be, but it is not exactly an established, going concern either: at best, it is a fledgling enterprise without fully audited assets and liabilities.

Under such circumstances, when confronted with an argument from premisses about dependence to the conclusion that set-theoretic ontological reductionism is false, should we not simply perform a Moorean shift and reject the conclusion out of hand?

[6] For an excellent treatment of the details, see Feferman (1989).
[7] For more on deference, see Maddy (1997).

If orthodox mathematics tells us that von Neumann reductionism is true, does this not itself give us reason to think that one of the premisses about dependence in the anti-reductionist argument must be false, even if we do not know which one? The anti-reductionist must confront this claim—the *challenge from orthodox practice*—squarely; if it succeeds, his entire project, however technically interesting it may be, will have little philosophical interest.

This challenge from orthodox practice would be compelling if set-theoretic reductionism were in fact one of the deliverances of standard mathematics. But it is not: appearences to the contrary are deceptive. We should start by noting that the authors of set theory textbooks do not claim that, in setting out the von Neumann and associated reductions, they have captured the genuine nature of natural numbers, real numbers, and the like by giving a correct analysis of an antecedently understood notion. Rather, they simply note that *the details of the construction make no mathematical difference*: definitions in set theory are good enough for mathematical purposes. Kenneth Kunen's formulation, in the most recent edition of his standard textbook, is exemplary:

> All mathematicians know that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. They may not know, or even care, what each real and complex number is as a set. They probably are aware that most (or all, depending on the exposition) of these $\subseteq$s aren't really $\subseteq$s after all, but are shorthands for embeddings.
>
> For example, we all know that $3 + 4i \in \mathbb{C} \setminus \mathbb{R}$ and $|3 + 4i| \in \mathbb{N}$. But what is the true identity of $3 + 4i$? What is its rank? This is usually irrelevant. It is important at some point to prove rigorously that it makes sense to compute with complex numbers. The simplest way of doing this is to define the complex plane $\mathbb{C}$ to be the set $\mathbb{R} \times \mathbb{R}$, so i denotes $(0,1)$ and $3 + 4i$ denotes $(3,4)$. Then $\mathbb{R}$ isn't a subset of $\mathbb{C}$ after all, but there is a natural embedding $x \mapsto \langle x, 0 \rangle$ from $\mathbb{R}$ into $\mathbb{C}$. [...]
>
> After learning this once, and verifying that the fields of real and complex numbers behave the way "you think they should," it is usually safe to forget

the details. [...]

There are many different expositions of these various number systems. (Kunen 2013: 80)

Kunen expressed this attitude more succinctly in an earlier edition of the textbook: "It is a philosophical quibble whether the elements of ω [i.e., ω_{VN}] are the real natural numbers (whatever that means). The important thing is that they satisfy the Peano postulates" (1980: 19).[8]

Unsurprisingly, a philosophical quibble is of more interest to philosophers of mathematics than to mathematicians. But where a question is of only peripheral interest to a scientific discipline, the claim that philosophers should defer to the opinions of practitioners of the discipline about that claim is correspondingly weakened. As such, *intertheoretic identity claims*—theses like $\mathsf{On} = \mathsf{On}_{VN}$, $\mathbb{N} = \mathbb{N}_{VN}$, $\mathbb{Q} = \{X : X \text{ is an equivalence class of } \mathbb{N}_{VN} \times \mathbb{N}_{VN} \text{ under } \sim, \text{ where } \langle a,b\rangle \sim \langle c,d\rangle \leftrightarrow ad = cb\}$—should not be counted as part of the "established, going concern" towards which deference is appropriate.

Instead of examining mathematicians' *obiter dicta* about intertheoretic identities, we should look at the ways in which working mathematicians actually integrate different areas of mathematics, and here the evidence is clear: for the working mathematician, mathematics is a disparate collection of subfields with varying subject matter, whose interconnections are not clear in advance but explored as the need arises, without any overarching commitment to ontological reduction. On this point, the characterization of mathematics given by Saunders Mac Lane (1986: 417) is particularly trenchant: "We cannot realistically constrain Mathematics to a single formal system; instead we view Mathematics as an elaborate tightly connected network of formal systems, axiom systems, rules, and connections." After devoting many pages to setting out and commenting on diagrams of the "nodes of that network—the various special 'subjects' into which Mathematics is (or may be) divided" (1986: 422), he concludes:

[8] Cf. Quine (1960b): 256–65.

"We have thus illustrated many subjects and branches of Mathematics, together with diagrams of the partial networks in which they appear. The full network of Mathematics is suggested thereby, but it is far too extensive and entangled with connections to be captured on any one page of this book" (1986: 428).

Although Mac Lane's resolute anti-realism about the objects of mathematics is philosophically unsatisfying, his description of the way mathematicians navigate the various subfields of the discipline is wholly persuasive. Mathematics, as actually studied, is a motley world. Mathematicians do real analysis starting from axioms for $\mathbb{R}$, complex analysis starting from axioms for $\mathbb{C}$, number theory starting from axioms for $\mathbb{N}$ (or simply from an intuitive grasp of the real numbers, complex numbers, or natural numbers, respectively), pure set theory starting from ZFC (or even from a fragment of naïve set theory, when they can be sure that what they need is capable of consistent reformulation in ZFC), group theory either by laying down the group axioms and then considering examples from other areas of mathematics or by making general claims about all groups, and so on. Moreover, when they wish to explore the interconnexions among these subfields, they usually either construct (possibly unintended) models for one theory out of the objects posited by another or develop relative interpretations—procedures which do not involve pronouncing on intertheoretic identity claims.

This aspect of mathematical practice is most patent in number theory and analysis. The classic treatise on number theory opens not with any characterization of the integers in terms of anything else, but simply with the statement that "[t]he numbers $\ldots,-3,-2,-1,0,1,2,\ldots$ are called the *rational integers*, or simply the *integers*" (Hardy and Wright 1960: 1). Similarly, a textbook on real analysis will likely begin not with a construction of real numbers in terms of Dedekind cuts or Cauchy sequences, but with a delimitation of the reals in terms of their most important structural characteristic: that they form an ordered field with the least-upper-bound property.[9] The details of the construction are secondary, and, as John Burgess has noted, those details *do not*

[9]See, e.g., Rudin (1976: 3–12).

form part of the content (in the narrow sense) of the subfield of real analysis itself:

> [D]ifferent textbooks introduce the real numbers in different ways, and while reviewers may find one approach *pedagogically* superior to another, no one considers any definition *mathematically* superior to any other, so long as that other still suffices for deducing the basic properties of the real numbers, those of a complete ordered field. Two analysts who wish to collaborate do not need to check whether they were both taught the same definition of "real number," as conceivably two algebraists may have to check whether they are both using the same definition of "ring" (since on some usages the definition involves having a multiplicative identity, while for others it does not). For it is only the properties of a complete ordered field that will be used in their collaboration, and never the definition of "real number."
>
> Thus while a textbook may offer a definition of the real field as composed of pairs of sets of rationals *à la Dedekind*, or alternatively as classes of sequences of rationals *à la Cantor*, and on the basis of its definition prove the theorem that the real field is a complete ordered field, only the theorem and not the definition has the status of being generally accepted by the mathematical community and used in subsequent mathematical work. (Burgess 2015: 152)

Considerations like these have impelled many mathematicians to some version of structuralism: they lie at the basis of Benacerraf's argument discussed in §1.3 below. But, before that question (which we can subdivide in a manner corresponding to the different varieties of structuralism) arises, we confront a preliminary one: if mathematics is the sort of complex network that Mac Lane suggests, its component parts existing in relative autonomy from one another, what is the position of set theory within that network?

Total pure set ontological reductionism offers a simple answer, but total pure set on-

tological reductionism goes well beyond what respect for mathematical practice commits us to. This does not itself provide an argument against reductionism, but it means that the answer to the question cannot simply be read off mathematical practice. Yet it is also implausible that set theory is nothing more than one field of mathematical inquiry among many, with no special status: even if we do not accept set-theoretical reductionism, we need to explain why it seems *prima facie* plausible; more generally, we need to account somehow for the privileged foundational role of set theory in contemporary mathematics.

Once the problem is posed in this way, an answer suggests itself straightforwardly: pure set theory has a preeminent role, but not because it is any part of the deliverances of orthodox mathematics that every mathematical object is a set; rather, it is a presupposition of the working mathematician's practice that every mathematical structure to be studied has an adequate *pure set-theoretical surrogate*—in other words, a structure *that can go proxy for the original structure for any mathematical purpose*, where 'mathematical purpose' is narrowly construed (*i.e.*, where it does not include metaphysical investigation of an intertheoretic identity thesis).

This surrogacy principle—the claim that working in set theory is always good enough for mathematical purposes—is amply confirmed by the last hundred years of mathematical practice. To reject it would be to engage in gross revisionism. In general terms, we should endorse:

(1.1.1) For every structure $\mathscr{S}$ studied in any mathematical discipline, there exists a structure $\mathscr{S}'$ within V such that $\mathscr{S}'$ is an adequate mathematical surrogate for $\mathscr{S}$.

At a minimum, $\mathscr{S}'$ is an adequate mathematical surrogate for $\mathscr{S}$ only if any *purely mathematical* inquiry that can be carried out concerning $\mathscr{S}$ can be carried out equally well using $\mathscr{S}'$. (If pure-set reductionism holds, then (1.1.1) will obtain trivially, since $\mathscr{S}$ is clearly an adequate surrogate for itself.)

Can we say anything further about the notion of mathematical surrogacy? One

thing is implicit in the observation, discussed above, that the precise details of (for example) different set-theoretic constructions of the real numbers do not matter for doing real analysis. The reason a structure $\mathbb{R}_{Ded}$ based on a Dedekind-cut construction and a structure $\mathbb{R}_{Cau}$ based on a Cauchy-sequence construction are adequate mathematical surrogates for the true real numbers $\mathbb{R}$ and for each other (again, leaving open the question whether $\mathbb{R}_{Ded} = \mathbb{R}$ or $\mathbb{R}_{Cau} = \mathbb{R}$) is that they are both isomorphic to $\mathbb{R}$ (and thus isomorphic to each other). The general principle is clear:

(1.1.2) If $\mathscr{S}$ is isomorphic to $\mathscr{S}'$, then $\mathscr{S}'$ is an adequate mathematical surrogate for $\mathscr{S}$.

It should be noted that (1.1.2) is not wholly free of ambiguity, for we have not specified how to cash out the notion of isomorphism: if it is formulated in set-theoretic terms, rather than using higher-order resources, it will not be applicable to proper-class-sized mathematical structures. And this is clearly inadequate to the cases we are interested in. Not only do we expect (for instance) On_{VN} to be a suitable surrogate for On, but restriction to set-sized structures rules out even trivial cases: V, for instance, is clearly an adequate mathematical surrogate for itself. In §§1.2 and 4.3, this will be discussed in more detail, but for the moment we shall assume a higher-order formulation has been fixed and continue to use the looser wording here.

Acceptance of (1.1.1) and (1.1.2) opens the door for a picture that respects the primacy of set theory without assuming total pure set ontological reductionism; by locating set theory in its proper position within the motley world of mathematics, we avoid the problem that "[l]arge parts of mathematical practice are distorted by the demand that all mathematical objects be realized as sets (as opposed to being isomorphic to sets)" (Barwise 1975: 8).

Thesis (1.1.2) has not been formulated as a *necessary* condition for adequate mathematical surrogacy. Someone who took a resolutely first-order view, for instance, might claim that elementary equivalence, rather than isomorphism, is sufficient. But this position has some unpleasant consequences: it is hard to stomach the idea that a col-

lection of pure sets forming a nonstandard model of true arithmetic is an adequate surrogate of $\mathbb{N}$. Even if we add an equicardinality requirement, it still seems grossly implausible that a countable nonstandard model of arithmetic is an adequate surrogate of $\mathbb{N}$. Moreover, when we apply (1.1.1) to set theory itself, embracing an elementary equivalence criterion of surrogacy makes it hard to avoid an unacceptable form of Skolemite scepticism.[10]

Thus, although we can accept that there is a coherent notion of *weak surrogacy* on which a structure $\mathscr{S}'$, elementarily equivalent but not isomorphic to $\mathscr{S}$, can at least sometimes count as a weak surrogate for $\mathscr{S}$, when we are working with a notion of surrogacy that requires adequacy for the purposes of *any* purely mathematical inquiry, there is little reason not to accept the strengthening of (1.1.2) to a biconditional:

(1.1.3) $\mathscr{S}'$ is an adequate mathematical surrogate for $\mathscr{S}$ if and only if $\mathscr{S}$ is isomorphic to $\mathscr{S}'$.

Given (1.1.3) and (1.1.1), we derive our first important ontological claim about the set-theoretical universe:

(1.1.4) For every structure $\mathscr{S}$ studied in any mathematical discipline, there exists a structure $\mathscr{S}'$ within V such that $\mathscr{S}'$ is isomorphic to $\mathscr{S}$.

How does this picture apply to the cases of the cardinals and the ordinals? We start from the observation that in relatively informal treatments the cardinals (and, somewhat less frequently, the ordinals) are often presented in a fashion wholly detached from the specifics of the von Neumann definition.

Consider, for example, Herbert Enderton's discussion of cardinals in his *Elements of Set Theory* (1977). After defining the natural numbers as $\mathbb{N}_{VN}$ and introducing a cardinality function card for *finite* sets, he writes:

> The number card A measures the size of a finite set A. We want "numbers" that similarly measure the size of infinite sets. Just what these "numbers"

[10] See Benacerraf (1985) on these issues.

are is not too crucial, any more than it was crucial just what set the number 2 was. The essential demand is that we define card A for arbitrary A in such a way that

$$\mathsf{card}\, A = \mathsf{card}\, B \text{ iff } A \approx B.$$

Now it turns out that there is no way of defining card A that is really simple. We therefore postpone until Chapter 7 the actual definition of the set card A. [...] We define a cardinal number to be something that is card A for some set A. [...] But card ω [*i.e.*, here, card $\mathbb{N}_{VN}$] is not a natural number [...]. Just what set card ω is will not be revealed until chapter 7. Meanwhile, we give it the name that Cantor gave it:

$$\mathsf{card}\, \omega = \aleph_0.$$

(Enderton 1977: 136–7)

Enderton then proceeds to develop a substantial amount of cardinal arithmetic, including a cardinal-theoretic version of Cantor's theorem, $\kappa \neq 2^\kappa$ (1977: 142), before Chapter 7 (where he introduces On_{VN} and then carries out the standard identification of cardinals with Card_{VN}). Further, he directs the reader that "[i]f you plan to omit Chapter 7, then regard card A as an additional primitive notion and the promise as an additional axiom" (1977: 137).

Enderton assumes that card A is a set for all A, but in a higher-order fragment, if the notion is taken as primitive, with

$$\mathsf{card}\, A = \mathsf{card}\, B \text{ iff } A \approx B \tag{1.1.5}$$

as its characteristic axiom, then there is no reason to require this. Of course, in this setting, (1.1.5) becomes the set-theoretic analogue of an abstraction principle—specifically, Hume's Principle, which will be discussed extensively in the coming chapters.

And in the earliest treatments of cardinals, the assumption that a cardinal is itself a set of any type is nowhere made. In Cantor's original presentations of the notion of *Mächtigkeit* (power) in the 'Beiträg zur Mannigfaltigkeitslehre' (1878) and *Grundlagen einer allgemeinen Mannigfaltigkeitslehre* (1883), which covers general set theory, no requirement is laid down that the *Mächtigkeiten* themselves be sets. We find merely a version of (1.1.5) (restricted to sets that can be well-ordered):

> When two well-defined multiplicities M and N can be completely and exactly placed in the same order, element for element (which, whenever it is possible in one way, can also happen in many other ways), we shall permit ourselves in what follows to say that these multiplicities *have the same power* or that they are *equivalent*.[11] (Cantor 1878: 119)

> Each well-defined multiplicity possesses a determinate power, such that the same power is ascribed to two multiplicities when they can, reciprocally, be exactly placed in the same order, element for element.[12] (Cantor 1883: 167)

In *Beiträge I*, in contrast, it is claimed that "the cardinal number $\overline{\overline{M}}$ is itself a determinate set composed of bare unities", but this is a consequence of the specific theory of cardinals as resulting from a double act of psychological abstraction endorsed by the later Cantor. On this account, a cardinal number exists "as an intellectual image or projection of the given set M in our mind"[13] (1895: 283); the theory plays no mathematical role in the theory of cardinals.[14]

Unsurprisingly, before von Neumann's 1923 paper, this approach to characterizing

[11] "Wenn zwei wohldefinierte Mannigfaltigkeiten M und N sich eindeutig und vollständig, Element für Element, einander zuordnen lassen (was, wenn es auf eine Art möglich ist, immer auch noch auf viele andere Weisen geschehen kann), so möge für das Folgende die Ausdruckweise gestattet sein, daß diese Mannigfaltigkeiten *gleiche Mächtigkeit* haben, oder auch, daß sie *äquivalent* sind." In the absence of contrary indication, all translations are my own.

[12] "Jeder wohldefinierten Menge kommt danach eine bestimmte Mächtigkeit zu, wobei zwei Mengen dieselbe Mächtigkeit zugeschrieben wird, wenn sie sich gegenseitig eindeutig, Element für Element, einander zuordnen lassen."

[13] "Da aus jedem einzelnen Elemente m, wenn man von seiner Beschaffenheit absieht, eine »Eins« wird, so ist die Kardinalzahl $\overline{\overline{M}}$ selbst eine bestimmte aus lauter Einsen zusammengesetzte Menge, die als intellektuelles Abbild oder Projektion der gegebenen Menge M in unserm Geiste Existenz hat." See Cantor (1915: 86) for Jourdain's translation.

[14] For an attempt to make sense of Cantor's views on abstraction, see Fine (1998).

the cardinals was common in set theory textbooks: in his *Einleitung in die Mengenlehre*, for instance, after equating finite cardinals with the natural numbers, Adolf Fraenkel writes simply that "one [...] introduces a designation for that characteristic which is shared by all *infinite* sets that are equivalent to one another and speaks of the *cardinal number* or *power* of infinite sets" (1919: 41).[15] When Fraenkel revised the work as *Abstract Set Theory*, many years later, he included an extensive discussion of the notion of cardinal number, distinguishing among various approaches, including both the von Neumann reduction and a direct appeal to the Cantorian definition:

> Also with regard to cardinals it is sufficient to give a *working definition* such as the following, mentioned for finite cardinals above:
>
> > (B) The sets S_1 and S_2 are said to have *equal cardinals* if they are equivalent, i.e. if $S_1 \sim S_2$. Otherwise the cardinals are called *different*.
>
> As may be expected in view of ordinary arithmetic and will be confirmed in the following sections, all relations between cardinals can be reduced, in addition to the membership relation for sets, to the equality and inequality of cardinals, hence in view of (B) to the equivalence and non-equivalence of sets [...]. Accordingly we need not fix what a cardinal *is* and may translate statements on cardinals into the language of sets.
>
> One may take an additional step by altogether avoiding the use of cardinals, restricting oneself to corresponding *sets* and their equivalence. For ordinary purposes of set theory and most of its applications this would cause considerable inconvenience, though no difficulty of principle, but in axiomatic set theory [...] this way proves to be practicable. (Fraenkel 1966: 59–60).

Fraenkel develops a full account of cardinal arithmetic, and a partial account of ordinal

[15] "Entsprechend führt man nun auch für das Gemeinsame, was allen untereinander äquivalenten *unendlichen* Mengen jeweils eigentümlich ist, eine Bezeichnung ein, und zwar spricht man auch hier von der *Kardinalzahl* oder auch von der *Mächtigkeit* unendlicher Mengen." I am indebted to Paolo Mancosu for pointing me to this aspect of Fraenkel's work.

arithmetic, before introducing the von Neumann definitions, which appear only in an aside (Fraenkel 1966: 198–99).

Throughout *Abstract Set Theory*, Fraenkel treats the *informal* theory of cardinals and ordinals as a partially autonomous area of inquiry, whose content can be developed without resolving intertheoretic identity questions. Although advanced modern textbooks (Drake 1974: 21–54; Jech 2002: 17–36) frequently introduce On_{VN} and Card_{VN} directly, without any discussion of informal motivations, the *unofficial practice* of set theorists—particularly when set theory is studied in its interaction with other branches of mathematics—corresponds more to Fraenkel's approach than to the 'official' presentation. All that is required in the background is that On_{VN} be isomorphic (under the appropriate ordering) to On—a straightforward application of surrogacy by isomorphism. Again, this is not itself evidence that the von Neumann reductions are false—but it is evidence that they are not imposed upon us by orthodox mathematics.

1.2 Mathematical Subfields, Assemblages, and Theories, and Kinds of Reduction

This picture of the motley world of mathematics is quite informal. In order to render it more precise, we need to set out a number of distinctions among ways of systematizing mathematical inquiry. We can start by modelling mathematical practice as a whole as a collection $\mathscr{M} = \langle \mathscr{F}_1, \ldots, \mathscr{F}_n \rangle$ of subfields, which develop more or less independently. A subfield here is a mere tag, specifying a subject matter by any means convenient to mathematicians: group theory, analytic topology, number theory, real analysis, set theory, and so on are subfields. We make no assumption that we can directly read anything about the kinds of mathematical objects off this division into subfields.

It is tempting to think of a subfield as a covering category for a mathematical theory, which can then be regimented directly: for instance, to the subfield *set theory* corresponds the true theory T_{Set} of sets, whatever it is. If that were the case, then cataloguing the resultant theories would provide a framework for questions about intertheoretic

identities: the framework would not itself give us answers to these questions, but it would make them precise. But the situation is more complex than this. First of all, there are two different notions of *theory* that need to be prised apart: theories as logicians consider them—collections of sentences in a specified language, whether interpreted or uninterpreted—should be distinguished from theories in a richer, informal sense. Let us temporarily call the former an *L-theory* ('*L*' for *logical*) and the latter an *R-theory* ('*R*' for *rich*).

We write $T, T_1, \ldots$, and variants for arbitrary L-theories, $\hat{T}, \hat{T}_1, \ldots$, and variants for arbitrary R-theories. In general, where $\hat{T}_A$ is a specific R-theory, we write T_A for the L-theory corresponding to it. We should take these L-theories themselves to be individuated more finely than is normally required in logic: an L-theory T can be represented as a triple $\langle \mathscr{L}_T, Ax_T, \vdash_T \rangle$, where $\mathscr{L}_T$ is a language, Ax_T an axiomatization, and $\vdash_T$ a consequence relation. No assumption is made that Ax_T is recursive; true arithmetic, for instance, counts as an L-theory (although, of course, one we grasp very imperfectly). We can think of an R-theory $\hat{T}$ as represented by the pair $\langle T, I_T \rangle$ of an L-theory T and a *general interpretation (G-interpretation)* I_T of T.

G-interpretations are not merely structures or classes of structures, as is the case in standard model theory. Rather, they allow us to represent a much fuller range of options. In particular, we admit G-interpretations falling into these categories:

1. *assertory*, divided into
    (a) the two categories of
        i. *assertory and uniquely interpreted*,
        ii. *assertory and multiply interpreted*,
    (b) and, independently, into the two categories of
        i. *existentially committing* and
        ii. *non-existentially committing (generalist)*;
2. *non-assertory*, divided into

(a) *fictionalist*,

(b) *merely instrumental*,

(c) and *algebraic*, including

  i. *algebraic and subject-unrestricted* and

  ii. *algebraic and subject-unrestricted*.[16]

In contrast to G-interpretations, we temporarily term interpretations in the more familiar sense *S-interpetations* ('*S*' for *semantic*). But even S-interpretations are not to be equated with set-theoretic models: a less rigid notion is needed, for we wish (as discussed in §0.2) to admit the possibility of absolutely unrestricted quantification. S-interpretations can be thought of as structures over classes, but once again, this is a *façon de parler*. We use $\mathfrak{I}$ and variants for S-interpretations, and $\mathscr{I}$ and variants for nonempty classes of S-interpretations; we write $\|\mathfrak{I}\|$ for the class of objects in the first-order domain of $\mathfrak{I}$. To indicate that an L-theory T or a sentence ϕ of $\mathscr{L}_T$ holds in $\mathfrak{I}$ (on whatever notion of truth-on-an-interpretation is appropriate), we write $\mathfrak{I} \vDash T$ or $\mathfrak{I} \vDash \phi$. Later on, where there is no danger of confusion, I shall revert to the ordinary overloading of terminology in our use of 'theory' and 'interpretation'.

I shall consider the various kinds of R-theories in detail below, but the most important initial distinction is between *assertory* (case 1) and *non-assertory* (case 2) R-theories. With each R-theory $\hat{T}$ is associated a certain speech act-type π, such that a mathematician who competently makes use of $\hat{T}$ performs a series of token acts of type π, having as their content either $\hat{T}$ as a whole or subcomponents thereof. With π is associated a *correctness condition*. In case (1), the speech act is simply assertion and the correctness condition is *accuracy*: a mathematician who uses an assertory R-theory in the ordinary way literally *asserts* the sentences that occur therein, and her doing so is *correct* just in case these sentences are *accurate*.

If $\hat{T} = \langle T, I_T \rangle$, where I_T is an assertory G-interpretation corresponding to a unique S-interpretation $\mathfrak{I}$, then the accuracy of $\hat{T}$ consists in nothing more than the truth of

[16] My terminology here draws on Shapiro (1997: 40–41), who speaks of 'nonalgebraic' and 'algebraic' axiom systems, and Hellman (2003: 135), who uses 'assertory' and 'formal'.

the L-theory T on $\mathfrak{I}$, so $\hat{T}$ (resp. an asserted sentence ϕ of $\mathscr{L}_T$) is correct just in case $\mathfrak{I} \vDash T$ (resp. $\mathfrak{I} \vDash \phi$). Under these circumstances, $\hat{T}$ is *uniquely interpreted* (case 2.a.i).

It is important to distinguish correctness conditions from liceity conditions: plausibly, the norm of assertion is knowledge, not truth (Williamson 2000b: 238–69), and a mathematician who asserts a sentence ϕ that she does not know does something illicit; nonetheless, if ϕ is true, her assertion is correct. If the G-interpretation I_T corresponds to a plurality $\mathscr{I}$ of S-interpretations, then $\hat{T}$ is *multiply interpreted* (case 2.a.ii). The accuracy conditions for a multiply interpreted R-theory are more complicated: in principle, a number of different approaches to the semantics are possible. It is most natural, however, to equate accuracy with truth on *every* $\mathfrak{I} \in \mathscr{I}$. This results in a supervaluationist semantics: $\hat{T}$ (resp. an asserted sentence ϕ of $\mathscr{L}_T$) is correct just in case, for every $\mathfrak{I} \in \mathscr{I}$, $\mathfrak{I} \vDash T$ (resp. for every $\mathfrak{I} \in \mathscr{I}$, $\mathfrak{I} \vDash \phi$).

Cutting across the distinction between uniquely interpreted and multiply interpreted assertory R-theories is that between *existentially committing* (1.ii.a) and *non-existentially committing* (1.ii.b) R-theories. There is substantial debate about how to make the notion of existential commitment precise in general (Rayo 2007; Krämer 2014: 33–134), but the informal notion is good enough for our purposes: R-theories which prove the existence of more than one object on their intended consequence relation, or prove any instance of the schema $\exists x\, \phi(x)$ not provable from logic alone, are existentially committing. In all the cases that we are interested in, a non-existentially committing assertory R-theory will be *generalist*: it will comprise a set of axioms of the form $\forall x(\phi(x) \to \psi(x))$ for a *characterizing formula* ϕ; in other words, the theory will tell us that *if there are objects satisfying* ϕ, then they are thus and so, but it will not itself guarantee that there are any objects satisfying ϕ. The correctness conditions for non-existentially committing assertory R-theories are the same as those for their existentially committing counterparts, but in the case of generalist R-theories it is useful to introduce an additional notion of *nonvacuous correctness*: a uniquely interpreted generalist assertory R-theory $\hat{T} = \langle T, I_T \rangle$, with the characterizing formula ϕ, where I_T corresponds to $\mathfrak{I}$, is nonvacuously correct just in case $\mathfrak{I} \vDash T$ and $\mathfrak{I} \vDash \exists x\, \phi(x)$ (and *mutatis*

mutandis for an asserted sentence in $\mathscr{L}_T$); likewise, a multiply interpreted generalist assertory R-theory $\hat{T} = \langle T, I_T \rangle$, with the characterizing formula ϕ, where I_T corresponds to $\mathscr{I}$, is nonvacuously correct just in case, for every $\mathfrak{I} \in \mathscr{I}$, $\mathfrak{I} \vDash T$ and $\mathfrak{I} \vDash \exists x \phi(x)$ (and *mutatis mutandis* for an asserted sentence in $\mathscr{L}_T$).

None of this is the case for *non-assertory* R-theories. Here the G-interpretation does not occur by way of a structure or class of structures, and the correctness condition has nothing to do with truth (or any similar accuracy-based notion, such as truth-in-a-structure on a class of structures). Moreover, the speech-act-type π characterizing the R-theory is not assertion: non-assertory R-theories are *linguistically displayed* using declarative sentences, but their users do not assert the contents of those sentences.

The simplest illustrative case is that of a fictionalist R-theory (case 2.a), since the characteristic type of speech act has obvious extramathematical analogues. When Gielgud, playing Hamlet, uttered the sentence 'Man delights not me', he did not make a false assertion, even if it was not the case that man did not delight Gielgud. Rather, Gielgud made no assertion: he engaged in the pretence of making an assertion, and merely to engage in the pretence of asserting something is not to assert anything at all. And the correctness condition for Gielgud's utterance was conformity with the fiction: it was correct because it is part of the fiction *Hamlet* that Hamlet, speaking to Rosencrantz and Guildenstern shortly after they arrive at Elsinore, asserts that man does not delight him (using the words 'Man delights not me'). In the case of an actor on stage, correctness is determined down to precise word choice, but this is not always the case: if two children playing cops and robbers have agreed to make believe a particular stick is a gun, then an utterance of 'I've got a gun' may be correct just in case the speaker is holding the stick, leaving a great deal of freedom to the players.[17] In each case, however, the key point is that there are well-defined correctness conditions for the relevant utterances *which do not have to do with truth*, even though the utterances take the form of declarative sentences.

Something wholly analogous applies to a fictionalist mathematical R-theory, which

[17] See generally Walton (1990) on such cases of make-believe.

has as its correctness condition conformity to a specified 'mathematical fiction'. Consider the case of a nominalist who believes that there are no real numbers but that real analysis can be simulated using spacetime points. If she does not wish to maintain that the sentences of real analysis are really false or have a misleading surface grammatical form, she can treat them as non-asserted sentences occurring in a fiction. On her account, a sentence ϕ of real analysis's being correct (in conformity with the relevant mathematical fiction) can then be cashed out in terms of the literal truth of a corresponding sentence ϕ^* concerning spacetime points (in the same way that the conformity of an actor's utterance with a theatrical fiction can be cashed out in terms of the presence of the corresponding sentence-type at the appropriate point in the script).[18]

Merely instrumental R-theories (case 2.b), the paradigm of which can be found in Hilbert's explanation of the "surpassingly important and fruitful method of ideal elements" (1926: 165), are somewhat similar.[19] A Euclidean geometer need not believe that the projective plane genuinely exists in order to make use of projective geometry as a means for proving theorems about the Euclidean plane: it suffices that two-dimensional projective geometry is a conservative extension of two-dimensional Euclidean geometry. Similarly, a logician studying Robinson arithmetic (Q) will frequently find it easier to work in arithmetic with Δ_0 induction ($\mathsf{I}\Delta_0$), drawing on the fact that $\mathsf{I}\Delta_0$ is interpretable in Q (Nelson 1986); a certain kind of strict finitist could take this course even if he does not in fact believe that induction is total with respect to Δ_0 formulas; in so doing, he would be making a merely instrumental use of $\mathsf{I}\Delta_0$, and his invocation of an axiom of $\mathsf{I}\Delta_0$ should certainly not be taken as an assertion. In cases like these, the G-interpretation for the R-theory $\hat{T}$ corresponding to the L-theory $T = \mathsf{I}\Delta_0$ or $T =$ projective geometry is degenerate: only the syntactic properties of T itself are

[18] This position is inspired by Field's (1980) treatment of analysis, but it is not intended to represent Field's own considered view. In his (1980) and (1989: 79), he endorses an error theory, but in (1989: 3) he seems to suggest that mathematical claims, as ordinarily understood, should be seen as containing a covert 'according to the fiction' operator. See also Yablo (2005: 98) for a version of pretence fictionalism closer to that discussed here.

[19] "Eine völlig andere, ganz eigenartige Deutung und prinzipielle Erfassung des Unendlichkeitsbegriffes lernen wir an der Hand der so überaus wichtigen und fruchtbaren Methode *der idealen Elemente* kennen." Cf. Hilbert (1967: 372) for the standard translation.

at issue. We can formulate correctness conditions for a merely instrumental R-theory as follows: a merely instrumental $\hat{T}_1$ is *P-correct relative to* a non-instrumental $\hat{T}_2$ just in case T_1 has some desirable proof-theoretic property P relative to T_2. Examples of such properties are relative interpretability, conservativeness, synonymy, and so on: we shall discuss these in more detail when we characterize ontological reduction as distinguished from other kinds of reduction. A merely instrumental $\hat{T}_1$ is *correct simpliciter* just in case it is P-correct relative to a non-instrumental $\hat{T}_2$ for some P and $\hat{T}_2$ is itself correct (according to its own correctness conditions).

It can be debated whether any subfield in $\mathscr{M}$ contains a genuinely instrumental or fictionalist R-theory $\hat{T}$ (as opposed to a $\hat{T}$ that mathematicians with highly restrictive ontological views might mistakenly view as merely instrumental or fictionalist), although I have included the categories for completeness. But the applicability of the other main category of non-assertory R-theories, *algebraic theories* (case 2.c), to mathematical practice is patent. The axioms in an algebraic R-theory are not asserted; instead, they provide something like a classificatory template. Consider the case of group theory; undoubtedly, this counts as a mathematical subfield. We ignore for the moment uses of singular terms such as 'the cyclic group of order 2', which raise complex issues, and start by considering the characteristic axioms (let us call them GrAx). GrAx are not equipped with a specific interpretation, but merely serve to specify conditions such that any structure whatsoever satisfying those conditions is a group. As Shapiro notes, "no one is troubled by the fact that the commutativity of multiplication is independent of the axioms of group theory. This is because, on all accounts, group theory is not about a single structure that is unique up to isomorphism" (1997: 40).

We could take proofs in group theory to be proofs implicitly or explicitly beginning with GrAx and terminating in sentences that would be regimented directly within the language of group theory: for example, on this approach, the proof that a group has a unique identity element would take the form of a valid proof of $\exists!x\forall y(x\circ y = y \wedge y\circ x = y)$ from GrAx. The conclusion $\exists!x\forall y(x\circ y = y \wedge y\circ x = y)$ is not asserted, for the axioms themselves are not asserted. This kind of algebraic theory is *subject-unrestricted* (2.c.i):

anything whatsoever obeying GrAx is classified as a group, regardless of its specific nature. The epistemic purchase of such an R-theory is indirect: although no premiss and no conclusion in the proof of the uniqueness of the identity element is itself the kind of thing that can be known (since, in the absence of a specific interpretation, sentences in the language of group theory do not express propositions), once we have understood the proof, we are in a position to know—and to assert—that, *if* $\mathscr{G}$ is a group, *then* it has a unique identity element. The correctness conditions for a subject-unrestricted algebraic R-theory are trivial, but we can introduce a notion of nontrivial correctness as we did in the case of generalist assertory R-theories: a subject-unrestricted algebraic R-theory $\hat{T}$ is nontrivially correct just in case there exists some structure $\mathscr{S}$ such that $\mathscr{S} \vDash T$. We can also conceive, however, of a *subject-restricted* algebraic theory (case 2.c.ii)—where the characteristic axioms are implicitly designed only to apply to structures made up of objects of a particular kind K: in such a case, an R-theory $\hat{T}$ is nontrivially correct just in case there exists some structure $\mathscr{S}$ *over objects of kind K such that* $\mathscr{S} \vDash T$.

The close relationship between algebraic R-theories and generalist assertory R-theories is clear; but this very fact makes it difficult to claim that group theory is unequivocally to be regimented as an algebraic R-theory. An alternative to the algebraic approach would simply be to take the conditional claims one can come to know on the basis of reasoning from the group axioms as the direct content of the R-theory: thus the proof in our example would be represented as starting from a sentence expressing 'if $\mathscr{G}$ is a group, then $\mathsf{GrA}(\mathscr{G})$' (where $\mathsf{GrA}(\mathscr{G})$ is an appropriate object-language rendering of the claim that $\mathscr{G}$ satisfies GrAx) and concluding with a sentence expressing 'if $\mathscr{G}$ is a group, *then* it has a unique identity element': the corresponding R-theory would be assertory, but not existentially committing, with the characterizing formula GrA.

In practice, we find algebraists moving back and forth between the two modes of presentation. It is in the generalist form, for example, that I.N. Herstein gives the proof that we are using as an example—the uniqueness of the identity element in a group—in his *Topics in Algebra* (1975: 33). But in the chapter on Boolean algebras in

another elementary textbook, Birkhoff and Mac Lane's *Survey of Modern Algebra*, we find the claim that "starting from the commutative, associative, and distributive laws, we can derive by induction [...] generalized distributive laws such as the following: $x \wedge (y_1 \vee \cdots \vee y_n) = (x \wedge y_1) \vee \cdots \vee (x \wedge y_n)$" (1977: 364) (where $\wedge$ and $\vee$ represent Boolean meet and join), which is naturally understood as a metatheoretic claim about what can be proved from the axioms for Boolean algebras in their unconditional form.[20]

Cases like this suggest a second problem with the simple account on which a theory (R-theory or L-theory) is directly associated with a subfield: the practice of mathematicians working in the subfield does not always neatly track a single kind of G-interpretation. It will be useful to introduce a term for the unsystematized body of mathematical discourse corresponding to a subfield: let us call it an *assemblage*. It is the assemblages that function as the 'nodes' of Mac Lane's network; even R-theories, although they are crucial to philosophical investigation, stand at one remove from mathematical practice. The full picture, then, is one on which every subfield $\mathscr{F}$ corresponds to an assemblage $\mathscr{A}_{\mathscr{F}}$: $\mathscr{A}_{\mathscr{F}}$ provides the raw data for rational reconstruction, the end result of which is one or more R-theories $\hat{T}_1(\mathscr{A}_{\mathscr{F}}), \hat{T}_2(\mathscr{A}_{\mathscr{F}}), \ldots$, which collectively represent the systematized content of $\mathscr{A}_{\mathscr{F}}$. To every such R-theory corresponds an L-theory, but only when the R-theory is assertory does the body of mathematical practice represented by $\mathscr{A}_{\mathscr{F}}$ involve an endorsement of the *truth* of the sentences of the L-theory (on some S-interpretation or S-interpretations).

In general, where $\mathscr{F}$ is a subfield of the sort traditionally considered part of 'algebra', with a focus on some specific putative mathematical kind K—group theory, the theory of rings, the theory of vector spaces, and so on—$\mathscr{A}_{\mathscr{F}}$ corresponds to at least a *pair* of R-theories: a *subject-unrestricted algebraic R-theory* $\hat{T}_1$ containing the (uninterpreted) characteristic axioms for structures of kind K, and a corresponding non-existentially committing assertory R-theory $\hat{T}_2$, consisting of the conditional claim 'if x is a K, then $\phi(x)$', where ϕ is a formulation of the characteristic axioms. Thus $\mathscr{A}_{\mathscr{F}_{Group}}$,

[20] Compare the presentation of simple results for rings in their more advanced *Algebra* (Mac Lane and Birkhoff 1988: 86). I am indebted to Timothy Williamson for emphasizing the heterogeneity found in texts such as these.

the assemblage corresponding to the subfield of group theory, contains at least the subject-unrestricted algebraic *R*-theory $\hat{T}_{Group-1}$, to which the L-theory whose axioms are GrAx correponds, and the matching generalist R-theory $\hat{T}_{Group-2}$.

It may also contain an assertory theory, however, if references to 'abstract groups' in claims like the following theorem (Birkhoff and Mac Lane 1977: 148) are taken as involving genuine singular terms: "The only abstract groups of order four are the cyclic group of that order and the four group." Officially, Birkhoff and Mac Lane gloss talk of 'abstract groups' as a way of talking about properties of all isomorphic groups: "The notion of isomorphism is technically valuable because it gives form to the recognition that the same abstract group-theoretic situation can arise in entirely different contexts" (1977: 138). But it is possible to give such claims an assertory reading.

This option can be motivated on the basis of *ante rem* structuralism (Shapiro 1997). Shapiro's views will be discussed at greater length in §4.5, but the basic picture is one on which our informal use of 'structure' corresponds to two distinct kinds of mathematicalia: *concrete structures* (C-structures) and *abstract structures* (A-structures), also termed 'pure structures' (Linnebo and Pettigrew 2014: 269), 'patterns' (Shapiro 1997: 84), or 'abstract forms' (Linnebo 2008: 75). Given any equivalence class $\mathscr{C}$ of structures (whether concrete or abstract) under isomorphism (subject to suitable restrictions to avoid contradiction), there exists an A-structure shared by every $\mathscr{S} \in \mathscr{C}$ and nothing else; holding fixed a given $\mathscr{S} \in \mathscr{C}$, we write $\$(\mathscr{S})$ for this structure. Moreover, if $\mathfrak{S}$ is the A-structure shared by every $\mathscr{S} \in \mathscr{C}$, then $\mathfrak{S}$ itself is a member of $\mathscr{C}$, instantiating the same isomorphism type: $\$(\mathfrak{S}) = \mathfrak{S}$.[21] If this is the correct view, then the assemblage $\mathscr{A}_{\mathscr{G}_{Group}}$ should also contain an existentially committing assertory R-theory $\hat{T}_{Group-3}$; assuming this theory is uniquely interpreted, $\|\mathfrak{I}\|$ will be the class of abstract structures for groups—i.e., the class $\{\$(\mathscr{G}) : \mathscr{G} \text{ is a group}\}$. For our purposes, however, we need not decide whether $\mathscr{A}_{\mathscr{G}_{Group}}$ corresponds to $\hat{T}_{Group-1}$, $\hat{T}_{Group-2}$, and $\hat{T}_{Group-3}$, or

[21] Linnebo and Pettigrew (2014: 272) term this the *Instantiation* constraint on *ante rem* structure. In addition to this paper, see also Linnebo (2008: 75–76) for discussion of appropriate structure abstraction principles, as well as §4.4. Here $ is functioning as a structure abstraction functor, although we leave the framework informal in comparison to the abstraction principles discussed later. For another theory of A-structures, see Ketland (MS).

just to the first two R-theories. Nor need we decide the corresponding question for the other subfields normally classed as 'algebra'.

The case of category theory, however, requires particular care. In some cases, philosophers proposing category-theoretic foundations have suggested that the elementary theory of the category of sets, ETCS (Lawvere 1964), furnishes an alternative to orthodox set theory. When applied for such a purpose, ETCS has to be interpreted as an assertory theory, as Øystein Linnebo and Richard Pettigrew have noted (2011: 233). There are serious objections, however, to this use of ETCS as a foundational framework (Linnebo and Pettigrew 2011; Burgess 2015: 177–89). Outside this particular project, category theory as a whole is best understood as a very general subject-unrestricted algebraic R-theory (Feferman 1967, 1977; Hellman 2003). The justification is beyond the scope of this work, but the basic principle at issue is that a non-algebraic reading of category theory would conflict with the wholesale lack of interest shown by category theorists in specifying intended models. In a representative passage discussed by Geoffrey Hellman, Steve Awodey, a proponent of category-theoretic foundations for mathematics, remarks that "a category is *anything* satisfying these axioms. The objects need not have 'elements', nor the morphisms be 'functions', although this is the case in some motivating examples" (Awodey 1996: 213; cf. Hellman 2003: 134–35). It is for this reason, according to Hellman, that the category theorist makes no effort to offer any kind of informal elucidation of 'morphism' of the type that a set theorist offers of 'element' when he lays out his account of the iterative hierarchy. In contrast to the general theory of categories, the theory of a particular 'concrete' category raises different issues: one approach is to see each such theory as a subject-restricted algebraic R-theory. For instance, the theory of the category of sets—outside the particular foundational context motivating ETCS—can be understood as an algebraic R-theory constrained by the requirement that it can only be applied to a domain of pure sets.

With this taxonomy in place, we can define various notions of reduction. The standard notions of *proof-theoretic reduction* are relatively unimportant for our purposes, although I include them for completeness. Proof-theoretic reductions consider

only L-theories, not R-theories. In our notation: given L-theories $T_1 = \langle \mathscr{L}_1, Ax_1, \vdash_1 \rangle$ and $T_2 = \langle \mathscr{L}_2, Ax_2, \vdash_2 \rangle$, where $\mathscr{L}_1 \subseteq \mathscr{L}_2$, T_2 is *conservative over* T_1 just in case, for all $\phi \in \mathsf{Fmla}(\mathscr{L}_1)$, $Ax_2 \vdash_2 \phi$ only if $Ax_1 \vdash_1 \phi$. A *function* $^\circ : \mathsf{Fmla}(\mathscr{L}_1) \to \mathsf{Fmla}(\mathscr{L}_2)$ is a *relative interpretation* just in case $^\circ$ preserves sentential connectives and quantifiers under relativization, and where ϕ is a sentence, $Ax_1 \vdash_1 \phi$ only if $Ax_2 \vdash_2 \phi^\circ$; if, where ϕ is a sentence, $Ax_1 \vdash_1 \phi$ just in case $Ax_2 \vdash_2 \phi^\circ$, then $^\circ$ is a *faithful relative interpretation*. T_2 is relatively interpretable (resp. faithfully relatively interpretable) in T_1 just in case there exists some relative interpretation (resp. faithful relative interpretation) $^\circ : \mathsf{Fmla}(\mathscr{L}_1) \to \mathsf{Fmla}(\mathscr{L}_2)$.[22]

Although proof-theoretic reductions are highly relevant for merely instrumental R-theories, they do little to illuminate the specifically metaphysical issues involved in other types of R-theories. Instead, the notion that interests us is *ontological reduction*. It was already developed informally in §1.1; here I set out the full space of options. Alexander Paseau (2009: 36–37) has offered a tripartite classification of *objectual*, *structural*, and *partial-denotation reductions* very similar to the one I shall use. Paseau formulates the notions of objectual and structural reduction with respect to natural numbers and sets; here a more general characterization, with subcategories, will be given. The notion of a 'partial-denotation reduction', however, requires a slight caveat. Drawing on Hartry Field (1974), Paseau characterizes the notion as follows:

> A partial-denotation reduction attempts to combine standard objectual semantics with the idea that arithmetical terms and predicates can stand for the entities they would stand for in the von Neumann objectual reduction and the Zermelo one ... —but only partially. A set-theoretic interpretation partially accords with arithmetical language if each term partially denotes the entity the interpretation assigns to it; and there are many such interpretations. Thus, the ω-sequence of numerals partially denotes several ω-sequences of sets. (Paseau 2009: 36, fn. omitted)

[22] See Feferman, Kreisel, and Orey (1962) for these definitions, and see Tarski, Mostowski, and Robinson (1953) on the topic of interpretability more generally.

This formulation, however, combines issues that should be kept separate: how an R-theory $\hat{T}$ handles sentences to which a unique value is not directly fixed by the semantic apparatus—whether it applies a principle of supervaluation over structures of one sort or another, treats them as defective, or takes a different approach—is independent from the question of the R-theory's ontological commitments. Field (1974: 212) chooses the supervaluationist approach, which is the most appealing. As in §0.1, I shall also focus on it; nonetheless, other options are possible.

Where $\hat{T}_1 = \langle T_1, \mathfrak{I}_1 \rangle$ and $\hat{T}_2 = \langle T_2, \mathfrak{I}_2 \rangle$ are uniquely interpreted assertory R-theories, we say that $\hat{T}_1$ is one–one ontologically reducible to $\hat{T}_2$ just in case $\|\mathfrak{I}_1\| \subseteq \|\mathfrak{I}_2\|$. Where $\hat{T}_1 = \langle T_1, \mathscr{I}_1 \rangle$ is a multiply interpreted assertory R-theory and $\hat{T}_2 = \langle T_2, \mathfrak{I}_2 \rangle$ is a uniquely interpreted assertory R-theory, then we say that $\hat{T}_1$ is many–one ontologically reducible to $\hat{T}_2$ just in case, for every $\mathfrak{I} \in \mathscr{I}_1$, $\|\mathfrak{I}\| \subseteq \|\mathfrak{I}_2\|$. Where $\hat{T}_A$ is one–one ontologically reducible to $\hat{T}_B$, we write $A \rightharpoonup B$ as shorthand; where $\hat{T}_A$ and $\hat{T}_B$ are mutually one–one ontologically reducible, we write $A \rightleftharpoons B$; where $\hat{T}_A$ is many–one ontologically reducible to $\hat{T}_B$, we write $A \overset{\text{mult}}{\rightharpoonup} B$. Thus, Paseau's paradigm case of a 'partial denotation reduction'—in which $\langle \mathbb{N}, < \rangle$ is ambiguously assigned to every ω-sequence in V—counts as a many–one ontological reduction on our account. I shall discuss this particular case at greater length below: it provides the best example of *multiple-reduction structuralism*.

We also define a notion of many–many reduction: when $\hat{T}_1 = \langle T_1, \mathscr{I}_1 \rangle$ and $\hat{T}_2 = \langle T_2, \mathscr{I}_2 \rangle$ are multiply-interpreted R-theories, $\hat{T}_1$ is many–many reducible to $\hat{T}_2$ just in case there is a natural mapping $f : \mathscr{I}_1 \to \mathscr{I}_2$ such that, for every $\mathfrak{I} \in \mathscr{I}_1$, $\|\mathfrak{I}\| \subseteq \|f(\mathfrak{I})\|$; where $\hat{T}_A$ is many–many ontologically reducible to $\hat{T}_B$, we write $A \overset{\rightarrow}{\rightharpoonup} B$.

For algebraic R-theories, the notion of ontological reduction does not make sense, but we can define *algebraic reduction* as follows: where $\hat{T}_1$ is a subject-unrestricted algebraic R-theory corresponding to the L-theory T_1 and $\hat{T}_2 = \langle T_2, \mathfrak{I}_2 \rangle$ is a uniquely interpreted assertory R-theory, then we say that $\hat{T}_1$ is *algebraically reducible* to $\hat{T}_2$ just in case, for some $A \subseteq \|\mathfrak{I}_2\|$, there is a structure $\mathscr{S}$ with domain A such that $\mathscr{S} \vDash T_1$. Likewise, where $\hat{T}_1$ is an algebraic R-theory restricted to objects of kind K corresponding to the L-theory T_1 and $\hat{T}_2$ is as before, then $\hat{T}_1$ is *algebraically reducible* to $\hat{T}_2$ just in case, for

some $A \subseteq \|\mathfrak{I}_2\|$, there is a structure $\mathscr{S}$ with domain A, where every object in A is of kind K, such that $\mathscr{S} \vDash T_1$.

Now everything is in place that is required to give a more developed characterization of the areas of mathematics that interest us. Mathematical practice as a whole, as I argued in the last section, is represented by a collection of subfields, each of which corresponds to an assemblage, but here we are concerned with only two: number theory ($\mathscr{F}_{Nat}$) and set theory ($\mathscr{F}_{Set}$). With the assemblages $\mathscr{A}_{\mathscr{F}_{Nat}}$ and $\mathscr{A}_{\mathscr{F}_{Set}}$ are associated at least five R-theories: $\hat{T}_{Nat}$, the theory of natural numbers; $\hat{T}_{PureSet}$, the theory of pure sets; $\hat{T}_{Set}$, the theory of all sets whatsoever; $\hat{T}_{Card}$, the theory of cardinals; and $\hat{T}_{Ord}$, the theory of ordinals. These are, of course, *informal* R-theories; we need not assume that we can directly characterize the corresponding L-theories, although, in some cases, we can provide at least minimal constraints: clearly, T_{Nat} extends PA; $T_{PureSet}$ extends ZFC; and T_{Set} extends ZFCU. (Here and in what follows, we use 'extension' to include improper extensions.)

As for $\hat{T}_{Card}$, the theory of cardinals; and $\hat{T}_{Ord}$, the theory of ordinals, these are not normally treated separately, for (as discussed in §1.1) these R-theories are part of the informal practice of set theory as a whole. We should not assume that we have access to a recursive axiomatization of these theories—although axiomatic theories of ordinals have been proposed (Takeuti 1954, 1965). Instead, we can think of *partial axiomatizations* for these theories as consisting of the ordinary deliverances of cardinal and ordinal arithmetic; T_{Card} and T_{Ord} extend these.

How should we understand these R-theories? *Prima facie*, each is an assertory R-theory; in particular, set theory and number theory are generally taken to be the paradigm cases of assertory, existentially committing R-theories. We shall first treat $\hat{T}_{Set}$, $\hat{T}_{PureSet}$, $\hat{T}_{Card}$, and $\hat{T}_{Ord}$. If any one of these is assertory, presumably the others are: it is one thing to maintain (for example) that ordinals are not sets, even though they can be adequately simulated by von Neumann ordinals; to maintain that such a simulation could take place even though one of the theories in question is not assertory would be highly implausible. So we should expect that, if appearances are deceptive

and any one of $\hat{T}_{Set}$, $\hat{T}_{PureSet}$, $\hat{T}_{Card}$, and $\hat{T}_{Ord}$ is not assertory or existentially committing, then so are all of these R-theories.

Eliminative structuralism is the reinterpretation of a *prima facie* existentially committing assertory R-theory as an algebraic or generalist R-theory in disguise.[23] An eliminative structuralist construal of an R-theory like $\hat{T}_{PureSet}$, however, faces serious difficulties. Suppose that $\hat{T}_{PureSet}$ is subject-unrestricted algebraic and $T_{PureSet}$ extends ZFC^2. Under these circumstances, $\hat{T}_{PureSet}$ will be nontrivially correct only if there is some structure $\mathscr{S}$ such that $\mathscr{S} \models \mathsf{ZFC}^2$; but (assuming higher-order quantifiers are given the standard interpretation) this can be the case only if there exist inaccessible-many objects. But our normal reasons for supposing that there are at least that many things in the universe lie precisely in the deliverances of set theory itself; we take the *sets themselves* to provide enough things to guarantee the existence of $\mathscr{S}$. If the ontological commitments of set theory are liquidated, however, this approach is unavailable. A similar argument applies to a generalist construal of $\hat{T}_{Set}$, where T_{Set} extends ZFCU^2. If, on the other hand, $T_{PureSet}$ is a first-order theory such as ZFC, then by the Löwenheim–Skolem theorem, if $\hat{T}_{PureSet}$ is subject-unrestricted algebraic, it will be nontrivially correct provided that there exist infinitely many things in the universe. But to take this approach is to accept *unintended* interpretations of the power set operation in ZFC—on which it does not truly capture every subset—as being on a par with intended ones. This is implausible for the same reasons that it is implausible to accept nonstandard models of arithmetic as being on a par with the standard one. A mathematician who is prepared to countenance a powerful set theory, whether or not she believes that it has a unique intended model, should not accept the sort of thoroughgoing relativism about the power set operation that arises when models of ZFC with fewer-than-inaccessible-many objects are treated as fully kosher; to do so undermines the reasons for accepting a strong set theory in the first place.

For these reasons, $\hat{T}_{Set}$, $\hat{T}_{PureSet}$, $\hat{T}_{Card}$, and $\hat{T}_{Ord}$ are most plausibly understood to

[23] For a discussion of eliminative structuralism in general, see Parsons (1990: 306–18). Dummett (1991: 296) terms the view 'hardheaded' structuralism.

be assertory, existentially committing R-theories.[24] But are they uniquely interpreted or multiply interpreted? As noted in §0.2, in the case of $\hat{T}_{PureSet}$, these two positions—*universism* and *pluralism*, respectively—have been the object of much debate. We write $\mathscr{I}_{PureSet}$ for the class of S-interpretations associated with $\hat{T}_{PureSet}$ (and likewise for the other R-theories): if $\mathscr{I}_{PureSet}$ has only one member, then universism is true. I shall not discuss extreme versions of pluralism, where $\mathscr{I}$ contains some $\mathfrak{I}$ of deviant cardinality (such as a countable model of ZFC) or where $\mathscr{I}$ contains distinct $\mathfrak{I}_1, \mathfrak{I}_2$ differing on their interpretation of power set.[25] There is, however, a variant of pluralism that suffers from none of these defects: *potentialism*. On the potentialist view, for any non-isomorphic $\mathfrak{I}_1, \mathfrak{I}_2 \in \mathscr{I}$ either $\mathfrak{I}_1$ is isomorphic to a proper rank-initial substructure of $\mathfrak{I}_2$ or vice versa: S-interpretations are *quasi-categorical*.[26] In the simplest version, due to Zermelo (1930), each inaccessible cardinality is represented in $\mathscr{I}$: every structure in $\mathscr{I}$ is (using, for convenience, an idiom more in keeping with a uniquely interpreted construal of set theory) isomorphic to V_κ for κ inaccessible and is thus a model of ZFC^2. The requirements on S-interpretations can be strengthened—for instance, requiring isomorphism to V_κ for κ measurable—while preserving quasi-categoricity.

If potentialism is true, then $\hat{T}_{Set}$, $\hat{T}_{PureSet}$, $\hat{T}_{Card}$, and $\hat{T}_{Ord}$ are presumably all multiply interpreted, and the only reasonable version of reductionism is many–many. For instance, suppose that $\mathscr{I}_{PureSet} = \{V_\kappa : \kappa \text{ is inaccessible}\}$ (again using universist language and eliding additional components of the signature for simplicity). In the uniquely interpreted case, von Neumann reductionism ($\mathsf{On} = \mathsf{On}_{VN}$) is a version of $Ord \rightharpoonup PureSet$: what is claimed is that the ordinals are a specific subclass On_{VN} of the specific class V. On the potentialist account, $\mathscr{I}_{Ord} = \{\{\alpha : \alpha \text{ is a von Neumann ordinal} \wedge \alpha < \kappa\} : \kappa \text{ is inaccessible}\}$ and the reduction $Ord \overrightarrow{\rightharpoonup} PureSet$ is witnessed by the

[24] For more on the difficulties with eliminative structuralism about set theory, see Parsons (1990: 332), Hellman (2001: 185–88). For an attempt to carry out the programme, see McGee (1997).

[25] For an example of extreme pluralism, see Hamkins (2012, 2015). There are important similarities with the work of Woodin (2001a, 2001b, 2011) (although Woodin wishes to restrict pluralism sufficiently to settle CH) and Arrigoni and Friedman (2013); see Antos, Friedman, Honzik, and Ternullo (2015) for discussion.

[26] See Isaacson (2011) and Rumfitt (2015: 308–17) for a discussion of the significance of quasi-categoricity. Hamkins and Linnebo (MS: §3.4) term the simplest view of this type 'Grothendieck-Zermelo potentialism'.

mapping $\{\alpha : \alpha \text{ is a von Neumann ordinal} \wedge \alpha < \kappa\} \mapsto V_\kappa$. If a supervaluationist semantics is accepted for both multiply-interpreted R-theories, the intertheoretic identity '$\mathsf{On} = \mathsf{On}_{VN}$' comes out supertrue; plausibly, then, this approach counts as a genuine form of von Neumann reductionism.

For the generality relativist, potentialism is a natural position. She accepts that, for any candidate quantifier domain $\mathscr{D}$, there exists a $\mathscr{D}^+$ extending it. For simplicity, consider the case of $\hat{T}_{PureSet}$: for any domain $\mathscr{D}$, let $\mathscr{V}$ be the restriction of $\mathscr{D}$ to pure sets. On a plausible potentialist account, the 'diagonalizing-out' procedure that leads from $\mathscr{V}$ to $\mathscr{V}^+$ corresponds to the transition from V_κ to V_λ for λ the least inaccessible greater than κ. In contrast, the generality absolutist will struggle to find a motivation for potentialism. Since an absolutely unrestricted first-order quantifier is available, the source of the multiplicity of interpretations must lie elsewhere. A potentialist account that is compatible with generality absolutism has been proposed by Gabriel Uzquiano (2015), but it has several disadvantages. Since the domain is fixed, whether or not an object is a set is interpretation-relative. Uzquiano formulates his theory with an object-language interpretational modal operator (where an interpretation $\mathfrak{I}_2$ is accessible from $\mathfrak{I}_1$ just in case $\mathfrak{I}_2$ extends $\mathfrak{I}_1$), yielding the result that $\exists x(\neg\mathsf{PureSet}\, x \wedge \Diamond\mathsf{PureSet}\, x)$ comes out true in every S-interpretation. Because the modal operator is interpretational, rather than metaphysical, this claim does not itself conflict with the highly intuitive thesis that, if it is metaphysically possible that something is a pure set, then it is metaphysically necessary that it is a pure set. But the combination is difficult to motivate. Implicit in Uzquiano's picture is the idea that the various $\mathfrak{I} \in \mathscr{I}_{PureSet}$ are on a par; but, against the background of absolutely unrestricted quantification, there exists a class $\mathfrak{V} = \bigcup\{\|\mathfrak{I}\| : \mathfrak{I} \in \mathscr{I}_{PureSet}\}$ on the basis of which we can construct an S-interpretation $\mathfrak{I}_\mathfrak{V}$ where the extension of PureSet coincides with all and only those things which are such that it is metaphysically possible for them to be pure sets. $\mathfrak{I}_\mathfrak{V}$ is not an element of $\mathscr{I}_{PureSet}$: this would violate the requirement of indefinite extensibility motivating the account. But $\mathfrak{I}_\mathfrak{V}$ is a far more *natural* S-interpretation of set theory than any $\mathfrak{I} \in \mathscr{I}_{PureSet}$: unlike them, it makes the metaphysical and inter-

pretational modalities coincide. Its only disadvantage is that there is no more capacious perspective from which $\mathfrak{V}$ is a set, rather than a proper class. For the generality relativist, this is a decisive objection. But the generality absolutist is already committed to S-interpretations that are not set-theoretic models: any assertory R-theory with absolutely unrestricted quantifiers (including, presumably, $\hat{T}_{Set}$) will have the universe as the domain of its S-interpretations.

From the generality-absolutist perspective adopted in this work, then, the most natural construal of $\hat{T}_{Set}$, $\hat{T}_{PureSet}$, $\hat{T}_{Card}$, and $\hat{T}_{Ord}$ is as uniquely interpreted assertory R-theories. Thus the candidate reductions that interest us—$Card \rightharpoonup PureSet$, $Card \rightharpoonup Set$, $Ord \rightharpoonup PureSet$, $Ord \rightharpoonup Set$, and $Card \rightharpoonup Ord$ (or, less plausibly, $Ord \rightharpoonup Card$) are simple one–one ontological reductions: they amount to the claims $\mathsf{Card} \subseteq \{x : \exists y\, y \in x\} \cup \{\emptyset\}$, $\mathsf{Card} \subseteq V$, $\mathsf{On} \subseteq \{x : \exists y\, y \in x\} \cup \{\emptyset\}$, $\mathsf{On} \subseteq V$, and $\mathsf{Card} \subseteq \mathsf{On}$ (or $\mathsf{On} \subseteq \mathsf{Card}$), respectively. (Technically, the domain of the S-interpretation asociated with $\hat{T}_{Set}$ is U, not $\{x : \exists y\, y \in x\} \cup \{\emptyset\}$, but since $\mathsf{Card} \subseteq U$ is trivial, we shall abuse notation and treat Set reduction claims as theses about ontological reduction to sets, not 'reduction' to objects in general.) As noted in §0.2, however, even if my arguments here are incorrect and $\hat{T}_{Set}$, $\hat{T}_{PureSet}$, $\hat{T}_{Card}$, and $\hat{T}_{Ord}$ are multiply interpreted, the antireductionist arguments in Chapters IV still tell against all the corresponding many–many reductionist claims, $Card \overrightarrow{\rightharpoonup} PureSet$, $Card \overrightarrow{\rightharpoonup} Set$, $Ord \overrightarrow{\rightharpoonup} PureSet$, $Ord \overrightarrow{\rightharpoonup} Set$, and $Card \overrightarrow{\rightharpoonup} Ord$. If *every* candidate reduction $\mathfrak{I}_1 \mapsto \mathfrak{I}_2$ can be ruled out for $\mathfrak{I}_1 \in \mathscr{I}_A$ and $\mathfrak{I}_1 \in \mathscr{I}_B$, then there can be no function $f : \mathscr{I}_A \to \mathscr{I}_B$ (let alone a natural one) witnessing a many–many reduction $A \overrightarrow{\rightharpoonup} B$.

The problem of understanding $\hat{T}_{Nat}$, however, is more difficult. Given a background ontology of sets, eliminative structuralist strategies do not face the problem of justifying the existence of enough objects for nontrivial correctness; they face no decisive objection on that front. But the eliminative approach is strongly revisionary with respect to our intuitions about natural numbers. *Prima facie*, there is strong pressure to take arithmetic as a uniquely interpreted assertory R-theory: the natural numbers seem the most familiar of mathematical objects. It is important to distinguish two intuitions

at work here: first, *subject-specificity*: it is not the case that just anything can count as a natural number; second, *individualism*: there are some *specific* objects that are given to us as natural numbers. The force of the two intuitions differs, as can be seen by considering the distinct responses that we have to the Caesar problem raised by Frege and neologicists and to Benacerraf's multiple-reducibility dilemma (discussed in §1.3 below). In §56 of *Grundlagen*, Frege rejected the adequacy of Hume's Principle alone as a foundation of arithmetic on the grounds of its failure to prove that, *e.g.*, $2_{Nat} \neq$ Julius Caesar.[27] He treated this as a decisive objection (Frege 1884: §56/1980: 68). Intuitively, absent some way of discharging the worry, he was right to do so: the Caesar problem must be solved, for an account that fails to *positively rule out* $2_{Nat} =$ Julius Caesar is missing something fundamental about the nature of the natural numbers. In contrast, it is not generally considered an absolute requirement that a theory positively rule out at least one of $2_{Nat} = \{\{\emptyset\}\}$ or $2_{Nat} = \{\emptyset, \{\emptyset\}\}$; most structuralist accounts do not. Structuralist treatments not involving concreta are at least minimally plausible candidates for a theory of the natural numbers, whereas anything that leaves open the possibility that Caesar could be a natural number is not. But if the subject-specific and individualist intuitions were on a par, ruling out $2_{Nat} =$ Julius Caesar and ruling out at least one of $2_{Nat} = \{\{\emptyset\}\}$ or $2_{Nat} = \{\emptyset, \{\emptyset\}\}$ would be equally pressing requirements.

The intuition of subject-specificity tells against construals of $\hat{T}_{Nat}$ as a subject-unrestricted algebraic R-theory or as a generalist R-theory. But it does not rule out the G-interpretation of $\hat{T}_{Nat}$ that Paseau (2009: 40) uses as an example: in our terms, on this construal, $\hat{T}_{Nat}$ is a multiply interpreted assertory R-theory, and $\mathscr{S}_{Nat}$ contains an S-interpretation corresponding to every ω-sequence of pure sets; there is thus a many–one ontological reduction of the form $Nat \overset{\text{mult}}{\rightharpoonup} PureSet$.[28] As discussed in §0.2, if a supervaluationist semantics is used, this results in a multiple-reduction structuralist framework in which *internal* arithmetical claims come out supertrue or superfalse,

[27] On the Caesar problem in general, see Hale and Wright (2001a). It might be more accurate to say that Frege rejected recourse to HP alone becaue of its failure to prove claims like $2_{Card} \neq$ Julius Caesar, but the objection is equally persuasive even without assuming $Card \rightharpoonup Nat$.

[28] It also does not rule out interpreting $\hat{T}_{Nat}$ as a *subject-restricted* algebraic R-theory, but these have no obvious advantages over the option discussed in the text and are harder to motivate.

but mixed claims come out neither supertrue nor superfalse. Because—unlike the Benacerrafian arguments discussed below—my arguments are effective against this reduction, as well as the simpler $Nat \rightharpoonup PureSet$ reduction, I need not provide a full rejoinder to this version of multiple-reduction structuralism. Nonetheless, it is worth briefly sketching the independent reasons to prefer a uniquely interpreted over a multiply interpreted construal of $\hat{T}_{Nat}$. The individualist intuition has at least some independent force: much of this force, I maintain, derives from the fact that it seems that we can have *de re* cognitive access to natural numbers as objects—that we can have, in Gödel's terms, "something like a perception" of them (Gödel 1964: 267). Gödel made this claim about the objects of set theory, but the case seems much clearer for the natural numbers. At least *prima facie*, it appears that not only can I form the *de dicto* belief that 263_{Nat} is prime, but I can also believe, *of* 263_{Nat}, that it is prime. Yet this is hard to square with the idea that terms like '263_{Nat}' fail to refer uniquely—and thus hard to square with the most natural semantics for multiple-reduction structuralism. I have no comprehensive account of the workings of these *de re* attitudes; but this ought not to cause us to be sceptical of their existence.[29] After all, we have no good bottom-up account of perception, but this is no impediment to theorizing about material objects on the basis of our perceptual knowledge of them. Likewise, it is not required that we have a developed account of mathematical epistemology in order to use our knowledge of mathematicalia as data in developing mathematical metaphysics.

1.3 Benacerraf's Argument

With this classification of types of R-theories and kinds of reduction in place, we can turn to the best-known antireductionist argument, due to Paul Benacerraf in 'What Numbers Could Not Be'. This argument is directed against one–one set-theoretic reductionism about the natural numbers. Benacerraf asks us to imagine two children, Ernie and Johnny, "sons of two militant logicists—children who have not been taught

[29]For one proposal, on which we gain knowledge about $\mathbb{N}$ by way of knowledge of *quasi-concrete* ω-sequences (such as types of Hilbert stroke numerals), see Parsons (2008: 152–85). On *de re* attitudes to numbers more generally, see Ackerman (1978).

in the vulgar (old-fashioned) way but for whom the pedagogical order of things has been the epistemological order" (Benacerraf 1965: 48). Each was taught set theory first and then arithmetic. Ernie's natural number terms were introduced using the definitions

$$0 = 0_{VN}, 1 = 1_{VN}, 2 = 2_{VN}, \ldots,$$

Johnny's using the definitions

$$0 = 0_{Z}, 1 = 1_{Z}, 2 = 2_{Z}, \ldots.$$

But this leads to an obvious conflict: Ernie and Johnny cannot both be right, yet it seems that "[e]ach was given an account of the meaning (and reference) of number words to which no exception could be taken and on the basis of which all that we know about or do with numbers could be explained" (Benacerraf 1965: 55).

From this fact, Benacerraf apparently draws a very strong conclusion—there are no natural numbers at all:

> I therefore argue, extending the argument that led to the conclusion that numbers could not be sets, that numbers could not be objects at all; for there is no more reason to identify any individual number with any one particular object than with any other (not already known to be a number). [...]
>
> Why so many interpretations of number theory are possible without any being uniquely singled out becomes obvious: there is no unique set of objects that are the numbers. Number theory is the elaboration of the properties of all structures of the order type of the numbers. The number words do not have single referents. Furthermore, the reason identification of numbers with objects works wholesale but fails utterly object by object is the fact that the theory is elaborating an abstract structure and not the properties of independent individuals, any one of which could be characterized

without reference to its relations to the rest. [...]

> Arithmetic is therefore the science that elaborates the abstract structure that all progressions have in common merely in virtue of being progressions. It is not a science concerned with particular objects-the numbers. The search for which independently identifiable particular objects the numbers really are (sets? Julius Caesars?) is a misguided one. (1965: 70–71)

Benacerraf seems to be proposing some version of eliminative structuralism for the natural numbers—i.e., that $\hat{T}_{Nat}$ is not a uniquely interpreted assertory R-theory, but an algebraic or generalist R-theory. Whether he himself intended this reading is contestable; in an essay written years later, he distances himself from the conclusion (Benacerraf 1998: 50–53).

As my main interest is not exegetical, I shall not enter into this debate; in any case, as Linda Wetzel (1989) and Mark Steiner (1975) have noted, the argument against $Nat \rightharpoonup Set$ (or, more generally, $K \rightharpoonup Set$ ontological reduction for a mathematical kind K) based on the existence of multiple candidate reductions is separable from Benacerraf's putative structuralist conclusion (which would be subject to the arguments against eliminative structuralism discussed at the end of the previous section). One can accept the former as persuasive while holding that *sui generis* antireductionism, rather than eliminative structuralism, is the correct position about objects of the kind in question. For instance, this view corresponds to the position taken about the $Ord \rightharpoonup Set$ reduction by Christopher Menzel:

> But are the ordinals really sets? A number of considerations would indicate that they are not. [...] As Benacerraf first pointed out, there are infinitely many well-ordered collections of sets which we could identify with the ordinal number sequence, but no arguments to show that any particular collection rather than any other is the genuine article. However, Benacerraf claims, if the number sequence really is some particular collection of sets, there ought to be such arguments. But there aren't, so it isn't. Now one

> might maintain steadfastly that nonetheless some particular collection just is the number sequence, arguments or not. This option, however, which for Benacerraf "borders on the absurd", only highlights the (or at least, a major) point of the argument, viz., the tenuousness of any philosophical commitment to the set/number identification: since there are no arguments available for singling out one particular collection over any other, the true believer's choice is no more than an act of faith, a groundless leap; his commitment is essentially unjustifiable. (Menzel 1986: 41–42; fns. omitted)

Clearly, if the negative portion of Benacerraf's argument is as compelling as Menzel thinks, then it will play a central role in any argument for antireductionism.

But is it compelling? In order to answer this question, we first need to clarify the structure of Benacerraf's argument. Much of 'What Numbers Could Not Be' suggests that the argument is essentially deductive, proceeding in the following form (for the $Nat \rightharpoonup Set$ case):

(1.3.1) If $Nat \rightharpoonup Set$, then $Nat \rightharpoonup PureSet$.

(1.3.2) If $Nat \rightharpoonup PureSet$, then $\mathbb{N} = A$ just in case A forms an ω-sequence under an appropriate ordering.[30]

(1.3.3) $\mathbb{N}_Z$ forms an ω-sequence under an appropriate ordering (*viz.*, $<_Z$).

(1.3.4) $\mathbb{N}_{VN}$ forms an ω-sequence under an appropriate ordering (*viz.*, $<_{VN}$).

(1.3.5) $\mathbb{N} = \mathbb{N}_Z = \mathbb{N}_{VN}$.

(1.3.6) $\mathbb{N}_Z \neq \mathbb{N}_{VN}$.

(1.3.7) Hence, $Nat \not\rightharpoonup Set$.

[30] Benacerraf also suggests that the ordering relation "must be at least recursive, and possibly even primitive recursive" (1965: 61). He retreats from this conclusion later, however (Benacerraf 1996), and it is difficult to render precise: on the most straightforward construal, since recursiveness is defined on sets of natural numbers, $<$ is *trivially* recursive regardless of whatever the numbers turn out to be. I thus ignore this condition. For an attempt to make out a broader notion of recursiveness on which it would be coherent, see Halbach and Horsten (2005).

Premisses (1.3.3), (1.3.4), and (1.3.6) are mathematically trivial. Premiss (1.3.1) is not explicit in Benacerraf's paper, but it is suggested by the fact that, in practice, he considers only pure set reductions. It will be discussed more fully in §4.1, but it is highly plausible. A set is a reasonable candidate to be a natural number only if all the objects in its transitive closure exist necessarily. But necessarily existing objects other than mathematicalia—such as universals—are, for various reasons having to do with their specific characteristics, unlikely to figure in any such set.

This leaves (1.3.2) as the crucial premiss, which Benacerraf apparently endorses:

> If the conclusions of the previous section are correct, then both boys have been given correct accounts of the numbers. [...] But if, as I think we agreed, the account of the previous section was correct-not only as far as it went but correct in that it contained conditions which were both necessary and *sufficient* for any correct account of the phenomena under discussion, then the fact that they disagree on which particular sets the numbers are is fatal to the view that each number is some particular set. (1965: 55–56)

But (1.3.2) is an exceptionally strong claim. The background motivation for it seems to be something like what Linnebo has termed the "incompleteness claim" (2008: 59)—the thesis, underpinning various versions of structuralism, that an object such as a natural number, if it exists at all, should be a *bare place in a structure*, without nontrivial nonstructural properties. There are serious problems with this thesis as thus formulated, however. First of all, it is hard to cash out what counts as a *bona fide* nonstructural property in such a way as to avoid obvious counterexamples such as abstractness, being a place in a structure, and the like. Secondly, and more importantly, if being F is a nontrivial nonstructural property, then so (presumably) is being non-F, and nothing can both fail to be F and fail to be non-F: the notion of an object with literally no nontrivial nonstructural properties is incoherent.[31]

[31] Some care would be required in formulating this objection in a supervaluationist framework, as the truth of $Fa \vee \neg Fa$ would not entail the truth of Fa or that of $\neg Fa$, but this complication is not relevant to the dialectical situation: here we are assuming $Nat \rightharpoonup Set$ (and thus that $\hat{T}_{Nat}$ is uniquely interpreted). See Parsons (1990: 334–35) for similar objections to the incompleteness claim.

A more plausible version of the claim would be that all nonstructural properties of any mathematical object are fully determined by its structural properties; this suffices to motivate the thesis that there cannot be two distinct mathematical structures differing only in properties other than those "necessary and *sufficient* for any correct account of the phenomena under discussion" (Benacerraf 1965: 55–56)—i.e., differing only in properties other than that of forming an ω-sequence under an appropriate ordering. In recent work, Shapiro has retreated to a position akin to this: "On the view in question, every property that 2 enjoys comes in virtue of its being that place in the natural number structure. This is because that is what 2 is" (2006: 121). But there is little incentive for anyone not already committed to a very strong version of structuralism to accept this thesis. In particular, it is wholly question-begging against a reductionist: if a $Nat \rightharpoonup Set$ reduction holds, then clearly the sets in question will have nonstructural properties, and there is no reason at all to think that those properties must be determined by the fact that the sets form an ω-sequence.

For this reason, it is worth reconstructing Benacerraf's argument in a more persuasive form—as an essentially *abductive* chain of reasoning, leading to the conclusion that antireductionism is the *best explanation* of the existence of multiple, apparently comparable candidate reductions. I formulate the argument generally, not just for $Nat \rightharpoonup Set$. Let $\mathscr{R}$ be the class of candidate pure-set reductions for K.[32] We have:

(1.3.8) If $K \rightharpoonup Set$, then $K \rightharpoonup PureSet$.

(1.3.9) If $K \rightharpoonup PureSet$, then some $f \in \mathscr{R}$ is the correct reduction for K.

(1.3.10) For no $f \in \mathscr{R}$ have we any reason to believe that f is the correct reduction for K.

(1.3.11) We have strong reason to believe that no $f \in \mathscr{R}$ is the correct reduction for

[32] Attempting precisely to formulate the notion of a 'candidate reduction' leads to complications, since a purely extensional characterization will not capture what is required for Benacerraf's argument: for instance, if we simply think of candidate pure-set reductions as injections from $\tilde{K}$ into V, then if pure-set reductionism is true, there is an obvious way of specifying the correct reduction: it is the identity mapping. But, of course, adducing this as an answer to Benacerraf's challenge is question-begging. Here it will suffice to leave the notion informal.

K.

(1.3.12) We have strong reason to believe $K \not\vdash PureSet$.

(1.3.13) We have strong reason to believe $K \not\vdash Set$.

Here (1.3.8), (1.3.9), and (1.3.10) are independent premisses; (1.3.11) is to be justified on the basis of (1.3.10), (1.3.12) on the basis of (1.3.11) and (1.3.9), and (1.3.13) on the basis of (1.3.12) and (1.3.8).

Premiss (1.3.9) holds by the definition of $\mathscr{R}$, and we can take it to be known. Given a reasonable epistemic closure assumption (*i.e.*, if we have strong reason to believe that ϕ and we know that ϕ only if ψ, then we have strong reason to believe that ψ), the fact that we know (1.3.9) will justify the step from (1.3.11) to (1.3.12); if we take (1.3.8) to be known, in virtue of arguments akin to those for (1.3.1) given above, the same will hold for the step from (1.3.12) to (1.3.13).

Both (1.3.10) and the passage from (1.3.10) to (1.3.11) are more dubious, however. Restricted to the choice between $\mathbb{N} = \mathbb{N}_Z$ and $\mathbb{N} = \mathbb{N}_{VN}$, the case is plausible, but in a more general setting we can often obtain some grounds to believe that one reduction is preferable to another. For instance, when considering $Ord \rightharpoonup Set$ reductions, Benacerraf's example does not generalize: $\mathbb{N}_Z$, unlike $\mathbb{N}_{VN}$, lacks an extension to limit ordinals, since there can be no infinitely descending $\in$-chains. Hence there is no such thing as a proper class On_Z directly extending $\mathbb{N}_Z$, and thus no way to develop a full version of *Zermelo reductionism* ($\mathsf{On} = \mathsf{On}_Z$) that would parallel the von Neumann reductionism of $\mathsf{On} = \mathsf{On}_{VN}$. In principle, at least, such considerations could give us clear reason to prefer one $f \in \mathscr{R}$ over others. I shall discuss this at greater length in §4.1, but for the moment we can grant Benacerraf the premiss (1.3.10).

The more important problem is that the passage from (1.3.10) to (1.3.11) depends on a strong assumption of epistemic optimism. Obviously, the step is not intended to be deductively valid; even as an inference to the best explanation, however, it relies on the assumption that our failure to discern the correct reduction is better explained by the nonexistence of such a reduction than by the limitations of our epistemic pow-

ers. And, for familiar reasons, this is a questionable thesis: semantic externalism, as Paseau has remarked, gives us reason to think that "speakers might not have transparent knowledge of the referents of their number terms" (Paseau 2009: 35).

Here an analogy with epistemicism about vagueness is instructive: given a class $\mathscr{C}$ of candidate cut-off points for a vague term T, the epistemicist will accept that for no $c \in \mathscr{C}$ have we any reason to believe that c is the cut-off point for T; nonetheless, she will deny that this gives us reason to believe that no $c \in \mathscr{C}$ is the cut-off point for T since, on her view, the facts that fix the fine details of the extension of T are not ones that we should expect to be scrutable. On one epistemicist picture, our *macroscopic* practices of language use—the ones to which we have access—fix only the fact that the cut-off point must lie in $\mathscr{C}$; *where* it lies in $\mathscr{C}$ is profoundly sensitive to *microscopic* facts—*e.g.*, details of the physical world, fleeting changes in mental states, and so on—of which we are unavoidably ignorant.[33] Likewise, the anti-Benacerrafian can claim that our use practices for standard terms that denote objects in $\|K\|$ fix only the fact that some reduction in $\mathscr{R}$ is correct; which one is correct is determined at a microscopic level beyond our ken. At the very least, these considerations undermine Menzel's claim that the anti-Benacerrafian must make a "groundless leap": externalism and a nuanced theory of the supervenience of meaning upon microscopic as well as macroscopic facts give principled reasons for resisting the step from (1.3.10) to (1.3.11).

This does not mean that (1.3.10) provides *no* evidential support for (1.3.11). In general, when we stand in no positive epistemic relation to $\phi(x)$ for any x, but we do stand in a positive epistemic relation to $\exists x\, \phi(x)$, some explanation for that fact is desirable. The most convincing such explanations are *local*—they turn on features particular to the case in question. For instance, I know that some of my beliefs are false, but there is no belief of mine such that I know it is false; this blindspot (in the sense of Sorensen 1988) can be explained by a combination of facts about my epistemic fallibility and facts about my practices of belief revision. Where no such local explanation is possible,

[33] See Williamson (1994: 201–9). The version of epistemicism presented here is somewhat cruder than Williamson's; for instance, I am ignoring higher-order vagueness as irrelevant to the main point.

we can only fall back on global explanations—as is the case of the metasemantic theory to which the epistemicist about vagueness, or the anti-Benacerrafian about reductionism, appeals. But proponents of epistemicism recognize that this reliance on a global explanation is a theoretical cost; their claim is that the virtues of the account outweight this cost. To persuasively rebut Benacerraf's argument, the defender of $Nat \rightharpoonup Set$ (or $K \rightharpoonup Set$ for a similarly situated mathematical kind K) needs to offer a similar account of the positive case for reductionism.

Finally, it should be reiterated that Benacerraf's argument does not tell at all against $K \overset{\text{mult}}{\rightharpoonup} Set$ reductions, because premiss (1.3.9) fails in such a setting: many–one reduction does not require the existence of any uniquely distinguished $f \in \mathscr{R}$. As was discussed in §0.2, this limitation does not afflict the antireductionist arguments that will be presented in Chapter IV.

CHAPTER II: TYPED FRAMEWORKS

As is the case in most works on the philosophy of mathematics, the main role of symbolic expressions in this text is to supplement the natural language (here, English) in which arguments are made. My interest is not primarily in formal languages—with exactly specified syntax, proof procedures, and model theories—for their own sake. Instead, for the most part, symbolic expressions will be treated as equipped with specific (non-model-theoretic) intended interpretations, their meanings fixed either through explicit definitions or through informal natural-language explications. When I use fully described formal languages with precise model theories, it is to establish relevant mathematical results. Transfer principles, whose status I shall discuss below, then allow these results to be lifted to the *informal* claims, set out in mathematical English, that are of primary importance.

The situation in this work is, however, somewhat more complicated than usual. Some of the devices I introduce to supplement mathematical English lack adequate natural-language paraphrases, because (on the intended interpretation) they represent a genuine increase in expressive power. In particular, this is the case with primitive quantifiers of arbitrarily high finite order, which I use extensively. These should *not* be understood as involving quantification over any kind of *objects*—instead, they represent direct generalization into predicate or sentence position, or into yet higher types lacking natural-language equivalents. As noted in the previous chapter, when singular terms (such as expressions for class) go proxy for these higher-order devices, they are mere *façons de parler*; here I explain how to eliminate them from the official idiom.

A set-theoretical model theory will be developed for expressions of these types, but this is merely heuristic: if the view taken here of these quantifiers is correct, a genuinely faithful semantics for them can only be given in a metalanguage that itself involves higher-order quantification. It is arguable that this was Frege's understanding of higher-order quantifiers; hints of it are also to be found in Prior (1971: 31–47) (although there primitive quantification into sentence position is not entirely distin-

guished from substitutional quantification with sentences as the substitution class). But it has received systematic defence only recently.[1]

In §2.1, I offer a short defence of the coherence of primitive higher-order quantification, an explanation of its motivation (which dovetails with the position of generality absolutism discussed in §§0.2 and 1.2), and an introduction to the role model theory will play. The specific functional type theory I employ is set out in §2.2, along with a basic extensional model theory. In §2.3, I develop less standard versions of non-extensional model theory. I give a possible-worlds based system in passing, but my main interest is in an extremely general hyperintensional model theory based on one developed by Reinhard Muskens (2007). Despite its elegance, Muskens's system has been little used by philosophers. A competing framework for hyperintensional type theory, the theory of applicative structures developed by Christoph Benzmüller, Chad Brown, and Michael Kohlhase (2004), has, however, been applied by Cian Dorr (2016) in his account of higher-type identity claims. I show that in a certain sense nothing is lost by use of Muskens's framework: possible worlds-based models and applicative structures can both be simulated within it. On my view, unlike its competitors, the Muskens framework can provide genuine insight into the *intended interpretation* of higher-order quantification (even though that interpretation is not model-theoretic). Finally, in §2.4, I define a few concepts that will be needed for the technical discussion of abstraction principles in §§4.2 and 6.2–3.

I use hyperintensional models in §6.3 in order to show that whether or not an abstraction principle avoids cardinality conflict can depend on difficult questions about the fineness of grain of reality. Related issues also arise, less programmatically, in my discussion of Agustín Rayo's version of abstractionism in §3.3. In principle, these models could also be used to formalize the grounding and essence operators used in §§5.1–2. This project will not be developed here, but I shall offer some methodological remarks in passing in Chapter V.

[1] See Rayo (2002, 2006), Rayo and Uzquiano (1999), Rayo and Yablo (2001), Rayo and Williamson (2003), Williamson (2000a, 2003, 2013: 235–61), Dorr (2016).

2.1 Higher-Order Quantification and Faithful Model Classes

Today, the Quinean view that higher-order logic is not legitimately logic (1968: 66–68) is far from an unquestioned orthodoxy. Nonetheless, there is little consensus as to how higher-order quantifiers should be understood: Do they range over individual objects of some sort—sets, properties, Fregean Concepts, or the like? Do they range over the same things that the first-order quantifiers range over, but taken plurally rather than individually? Or should they be understood in some other way—as not having domains of *things* at all? We shall term these families of positions *singularism*, *pluralism*, and *primitivism* or *anti-objectualism*, respectively. My position in this work is primitivist.

There are three main motivations for rejecting singularism: first, as pointed out by George Boolos (1984: 64–67), objectual semantics cannot deal with higher-order quantification in full generality, because of the inconsistency of objectual versions of unrestricted higher-order comprehension schemata. This is clearest in the case of set-theoretical objectual semantics. The condition of being non-self-membered makes perfect sense; if the Axiom of Foundation is true, then every set is non-self-membered. Thus it seems that it should make sense that there is a second-order condition which something falls under if and only if it is non-self-membered; unrestricted second-order comprehension accords with this claim, for the schema $\exists X \forall x (Xx \leftrightarrow \phi(x))$ has $\exists X \forall x (Xx \leftrightarrow x \notin x)$ as an instance. But if higher-order quantification is to be understood in set-theoretical terms, then the comprehension principle ultimately amounts to $\exists y (\mathsf{Set}\, y \wedge \forall x (x \in y \leftrightarrow \phi(x)))$ (assuming a homophonic translation of the other expressions); of course, the relevant instance $\exists y (\mathsf{Set}\, y \wedge \forall x (x \in y \leftrightarrow x \notin x))$ is inconsistent. The same problem arises if one interprets second-order quantifiers as ranging over universals or properties (understood as objects): instead of $\in$, we use a two-place η predicate, with $x \eta y$ pronounced 'x instantiates the property y' or 'x participates in the universal y': the schema $\exists y (\mathsf{Property}\, y \wedge \forall x (x \eta y \leftrightarrow \phi(x)))$ is inconsistent by the same Russellian argument. The argument is completely general: for any relation R between objects figuring in a putative semantics for higher-order quantification, the

schema $\exists y \forall x(Rxy \leftrightarrow \phi)$ is inadmissible if substituends for ϕ are not restricted, for $\exists y \forall x(Rxy \leftrightarrow \neg Rxx)$ is refutable by logic alone.

This argument can be resisted by predicativists and generality-relativists—but each of these positions amounts, in effect, to denying full second-order comprehension. The predicativist who holds a property-theoretic interpretation of the second-order quantifier, for instance, will accept only those instances of $\exists y(\mathsf{Property}\, y \wedge \forall x(x \eta y \leftrightarrow \phi(x)))$ where ϕ does not contain η and perhaps meets other syntactic restrictions. The set-theoretic generality-relativist will overtly endorse $\exists y(\mathsf{Set}\, y \wedge \forall x(x \in y \leftrightarrow x \notin x)$ but insist that the outer existential quantifier be treated as having a more capacious domain than the inner quantifier; in effect, he accepts only the weaker $\exists^{+} y(\mathsf{Set}\, y \wedge \forall x(x \in y \leftrightarrow x \notin x))$. As noted in §0.2, I shall not consider these approaches.

Moreover, on the assumption of *first-order* generality absolutism, even if comprehension is pared down sufficiently to avoid contradiction, objectual construals of the second-order quantifier vitiate one of the key expressive advantages of second-order logic: By Tarski's theorem, given reasonable background assumptions, the satisfaction predicate for a first-order language $\mathscr{L}$ over a base theory T strong enough to code up its own syntax cannot be defined in T. If we extend $\mathscr{L}$ to a second-order language $\mathscr{L}^2$ with full comprehension, without altering any of the nonlogical axioms of T, we can use the second-order quantifier to define a satisfaction predicate for the first-order fragment—as was already noted by Tarski (1935) in his earliest treatment of truth definitions. But if the intended interpretation of the first-order quantifier is absolutely unrestricted, then this technique will be unavailing if the only understanding of higher-order quantification on offer is an objectual one: if the second-order quantifier ranges over objects, then those objects must also fall within the range of an absolutely unrestricted first-order quantifier; but a second-order quantifier that is a mere restriction of the first-order quantifier does not provide the additional expressive power required in order to give an explicit semantics for the first-order fragment of the language under consideration.

In any case, there is another argument against objectualism of a less technical na-

ture, which cannot be resisted by retreating to predicativism or generality relativism. If, in keeping with the Quinean tradition, we do not believe that the truth of the sentence 'the Arch of Constantine is grey' requires the existence of any object (such as the universal *greyness* or the set of grey things) other than the Arch of Constantine, then we should not think that the mere use of a variable rather than a constant in predicate position changes the situation. A variable of a given syntactic category should be a bearer of commitment to objects if and only if a constant of the same category is. Just as both proper names and first-order variables bear commitment to objects, it should be the case that predicate variables and predicate constants both carry objectual commitment or neither does; given nominalism about properties-as-objects, predicate constants do not, so neither do predicate variables. Moreover, the mere fact that a predicate variable is bound rather than free does not alter the situation: variable-binding operators, such as quantifiers or lambdas, simply carry into the object language what already occurs in the metalanguage when we permit an open formula to be evaluated relative to every possible assignment.

Moreover, the full strength of nominalism about properties-as-objects is not required for this argument. Given reasonable background assumptions, it is guaranteed to us by logic alone that, for at least *some* open formulas, there is no universal or set that is required for the truth of claims involving that formula, since naïve property comprehension and naïve set comprehension are inconsistent. Call any such formula *locally nominalist*. Plausibly, when ϕ is locally nominalist and thus carries no objectual commitment of its own, the variable X also carries no objectual commitment on any assignment on which its semantic value is the same as that of ϕ. But then, assuming that some form of generalization on ϕ is available that licenses the inference from $\phi(a)$ to $\exists X\, Xa$, there exists a kind of higher-order quantification that we can make sense of that carries no additional objectual commitment beyond that involved in the values of the first-order variables. And there is no reason why we should not simply *stipulate* that we shall avail ourselves of this resource in interpreting *all* regimented higher-order quantification (even if it should happen that, in cases where the natural-language analogues

of regimented predicates are not locally nominalist, the semantics most faithful to our ordinary usage would involve commitment to universals or sets).

These arguments lead to a compelling case against the singularist understanding of higher-order quantification. They do not resolve the problem of interpreting the higher-order quantifier, however, for both pluralism and primitivism represent viable alternatives to singularism. On the *pluralist* proposal, developed by Boolos (1984, 1985), the value of a second-order variable is not *a thing*, but *some things*. We can call those things a 'plurality', but it must be understood that this singular noun—like Russell's 'class as many' (1903: §70)—is an abuse of language. The plural conception clearly satisfies the minimal adequacy requirements for singulary second-order logic in the presence of absolutely unrestricted first order quantification. There is no difficulty in understanding a higher-order term such as $(\lambda x.x \notin x)$ on the plural conception: there clearly *are some non-self-membered sets*, and that is all that is needed for the plural interpretation. But, as Williamson points out (2003: 456), plurals do not naturally generalize to provide interpretations of polyadic second-order quantification or functional quantification, and the nonexistence of a null plurality—the fact that, if there are some Fs, there is at least one F—requires some inelegant technical complications. A partial solution to these problems is provided by Rayo (2006), who has shown how to extend the plural conception to arbitrarily high finite order, at the price of some estrangement from natural-language intuitions.

The central difficulty with pluralism, however, is that plurals do not always behave as one would expect were plural quantification true quantification into predicate position. Plurals, as Williamson remarks (2003: 456–57), are *rigid*: of any things, whatever they are, it is the case that they could not have been any other things than they are. In contrast, an example such as

(2.1.1) Alcibiades is brave but it is metaphysically possible for Alcibiades not to have been brave,

seems to license an inference to

(2.1.2) $\exists X(X(\text{Alcibiades}) \wedge \Diamond\neg X(\text{Alcibiades}))$

by the second-order version of existential generalization.

The need to account for this distinction gives support to the alternative, primitivist reading of higher-order quantification. According to primitivism, the second-order quantifiers do not range over any *things* at all—neither individual things, nor pluralities—but are simply to be taken as primitive; there is no metalanguage not itself involving higher-order quantification that we can use to explain them. Nonetheless, *speaking loosely*, we can say that the quantifiers range over Fregean 'Concepts', where Concepts are not taken to be objects. This mode of expression is, as Frege said, to be taken "*cum grano salis*" (1914: 259/ 1979: 240); Williamson terms it the "crude Fregean account" and explains its limitations thus:

> If generality-absolutists respond by denying that concepts are things, they cheat by betraying the spirit of their original explanation. Rather, they should insist that the very use of the noun 'concept' makes the crude Fregean account inaccurate. [...] Perhaps no reading in a natural language of quantification into predicate position is wholly satisfactory. If so, that does not show that something is wrong with quantification into predicate position, for it may reflect an expressive inadequacy in natural languages. We may have to learn second-order languages by the direct method, not by translating them into a language with which we are already familiar. (Williamson 2003: 458–59)

In contrast, the *refined* Fregean account—the position I term 'primitivism'—explains second-order non-plural quantification by a satisfaction theory that appeals to third-order non-plural quantification, and so on. Here and throughout I shall make use of the sort of loose talk that Williamson describes. The official semantics for nth-order quantification is to be understood as formulated in an $n+1$th-order metalanguage, but I use the word 'Concept', translating Frege's *Begriff*, to go proxy for the 'values' of second-order variables, 'Superconcept' for the third-order case, and so on. I occasion-

ally use 'Relation' and 'Superrelation' in the polyadic case. The term 'entity' serves as a placeholder to generalize over quantification at all finite orders: officially, claims about 'entities' are abbreviations of schemata in a metalanguage admitting quantification at all finite orders (and, if neeeded, disjunctions and conjunctions of denumerable length). Likewise, as noted, talk of proper classes is to be understood in higher-order terms: 'there exists a class such that ... ' and 'there exists a Concept such that ... ' receive the same official explication.

How is it that we can develop competence with these logical resources? What does the "direct method" of learning higher-order languages, of which Williamson speaks, amount to? Although I have no complete account to offer of how we can develop the ability to use novel resources that represent a real increase in expressive power, it seems that there are two key procedures at work. The first is generalization from a basis in natural language. We understand what it is to use a sentence composed of a proper name and a predicate constant; we understand the device of replacing a proper name with a first-order variable and then generalizing on it; we then recombine these ingredients—making use of our access to something like a cross-type version of Gareth Evans's generality constraint (1982: 104)—to obtain the idea of replacing the predicate constant with a predicate variable and then generalizing on it. We recognize the possibility of introducing predicates of predicates, and we repeat the process to obtain third-order quantification; and so on, for every finite order. (Whether we can continue this process into the transfinite is a more complicated issue; it will be discussed in passing in §2.2.) At every level, our understanding of what generalization requires gives us a guide to the inferential constraints on higher-order quantifiers: for example, the parallel between the introduction rule for the first-order existential quantifier in a natural deduction framework ($\frac{\phi(t)}{\exists x \phi(x)}$) and what we expect of the second-order quantifier ($\frac{\phi(t)}{\exists X Xt}$) provides an informal justification for the full comprehension schema $\exists X \forall x (Xx \leftrightarrow \phi(x))$. An additional justification is provided by considering plurals: at a minimum, the primitivist reading should not be *expressively impoverished* relative to

the plural reading. Clearly, however, for any predicate applying to some things, there 'is' a plurality (speaking loosely) of all and only those things. Taken together, these considerations give us reason to reject predicativist restrictions on higher-order logic. Moreover, considering the case of plurals can take us further: even if we have no predicate in our language that applies to *these* things and no others, nonetheless these things are some things: there 'is' a plurality of these things. If the primitivist reading is to be as strong as the plural reading, then we should expect that, for any things, there is a Concept that applies to all and only those things, whether or not we have an expression for it in any language we can speak. Extending this principle up the hierarchy, we have the result that higher-order entities are 'abundant' (in the sense of Lewis (1983: 12)), not mere "shadow[s] cast on [...] particular[s] by predicates" (Armstrong 1978: 1.13).

The second procedure is heuristic reasoning from set-theoretic models. One of the central limitations to primitivism about higher-order logic is that the *intended* interpretation is not directly mathematically tractable, for on the primitivist account the quantifiers do not range over objects, and mathematical structures are collections of objects. Nonetheless, we can appeal to an analogue of the surrogacy theses discussed in §1.1: the overall success of set theory as a representational framework gives us reason to expect that the intended interpretation will have some sort of representational surrogate over set-theoretic models. The surrogacy relation will not ordinarily be isomorphism, as was the case for the structures under consideration in §1.1, since (assuming that for any objects whatsoever, there is a second-order entity applying to all and only those objects) a higher-order version of Cantor's theorem establishes that there are more second-order entities than objects (and likewise more third-order entities than second-order, and so on). Nonetheless, we should expect to be able to find models which meet weaker requirements of *faithfulness to logical truth* and *faithfulness to logical consequence*.

Specifically, we can formulate the following surrogacy claims: Let $\mathscr{L}$ be a higher-order language whose intended interpretation is not model-theoretic. For the moment, we take as given the notions of *informal logical truth on the intended interpretation* and

informal logical consequence on the intended interpretation; we discuss below how to formulate these more precisely.

(2.1.3) There exists a natural class of set-theoretic structures $\mathscr{M}$ and a truth-in-a-model relation $\models$ holding between sentences of $\mathscr{L}$ and structures in $\mathscr{M}$ such that, for all $\phi \in \mathsf{Sent}(\mathscr{L})$, ϕ is logically true on the intended interpretation just in case, for every $\mathfrak{M} \in \mathscr{M}$, $\mathfrak{M} \models \phi$.

(2.1.4) There exists a natural class of set-theoretic structures $\mathscr{M}$ and a truth-in-a-model relation $\models$ holding between sentences of $\mathscr{L}$ and structures in $\mathscr{M}$ such that, for all $\phi \in \mathsf{Sent}(\mathscr{L})$ and $\Gamma \subseteq \mathsf{Sent}(\mathscr{L})$, ϕ is a logical consequence of Γ on the intended interpretation just in case, for every $\mathfrak{M} \in \mathscr{M}$, if $\mathfrak{M} \models \psi$ for all $\psi \in \Gamma$, then $\mathfrak{M} \models \phi$.

We say that a model class $\mathscr{M}$ is *(strongly) faithful to logical truth* (relative to $\models$) if it meets the condition in (2.1.3) and *(strongly) faithful to logical consequence* (relative to $\models$) if it meets the condition in (2.1.4). (We shall introduce another version, weak faithfulness, below.)

The general structure of faithfulness theses has not received much discussion; nonetheless, two specific faithfulness theses, due to Georg Kreisel (1967), have been accorded significant attention. Kreisel uses Val for intuitive validity (our 'logical truth on the intended interpretation') and V for validity "in all set-theoretical structures" (1967: 152) (here, truth in every standard extensional model of higher-order logic). He glosses the difference as follows:

> The intuitive meaning of Val differs from that of V in one particular: $V\alpha$ (merely) asserts that α is true in all structures in the cumulative hierarchy, i.e., in all sets in the precise sense of *set* above, while $Val\,\alpha$ asserts that α is true in *all* structures [...]. (Kreisel 1967: 153)

If 'structures' are intended to be objects, then this is a gloss on Val that the primitivist cannot accept, for it amounts to countenancing something like proper classes alongside

sets, rather than understanding class talk as a mere *façon de parler.* But there are other ways the primitivist can explain informal logical truth on the intended interpretation. For instance, we can invoke what Vann McGee terms *Tarski's thesis* (2006: 100–3), the equivalence of informal logical truth on the intended interpretation of a sentence ϕ with truth of the sentence ϕ^* that results from ϕ as follows: Let ϕ^F be the result of relativizing every quantifier in ϕ to the predicate term F.[2] Let $\phi^\dagger$ be the result of replacing every nonlogical constant with a fresh variable and then taking the universal closure; let ϕ^* be $\exists X(\exists x\, Xx \wedge (\phi^\dagger)^X)$. In essence, ϕ^* is a modification of Tarski's original (1936) proposal for fixed-domain logical truth (Williamson's (2013: 93) "metaphysical universality") to allow for variable-sized (nonempty) domains; accordingly, we term ϕ^* the *variable-domain Tarskification* of ϕ. We term $\phi^\dagger$ itself the *fixed-domain Tarskification* of ϕ.[3]

With Tarski's thesis in place, we can formulate (2.1.3) for an nth-order language $\mathscr{L}^n$ in a sufficiently strong $n+1$th set theory (in $\mathscr{L}^{n+1}_{[\in]}$, a language that results from introducing $n+1$th-order variables and adding $\in$ to the signature, if needed).[4] Logical consequence is more difficult, but if the cardinality of $\mathsf{Sent}(\mathscr{L}^n)$ is κ, then Tarski's Thesis can be used to formulate a version of (2.1.4) in an appropriate set theory couched in the infinitary language $\mathscr{L}^{n+1}_{\kappa^+\omega[\in]}$ (admitting conjunctions of length κ) by moving from ϕ's being a logical consequence of Γ to the sentence $\bigwedge\Gamma \to \phi$'s being a logical truth. In any case, these extensions will not be needed here; an informal account of logical truth

[2] Higher-order quantifiers must be relativized appropriately: although $(\forall x\,\phi)^F$ becomes $\forall x\,(Fx \to \phi^F)$ and $(\exists x\,\phi)^F$ becomes $\exists x\,(Fx \wedge \phi^F)$, $(\forall X\,\phi)^F$ becomes $\forall X\,(\forall z(Xx \to Fx) \to \phi^F)$ and $(\exists X\,\phi)^F$ becomes $\exists X\,(\forall z(Xx \to Fx) \wedge \phi^F)$, and so on at higher order.

[3] The process, it should be noted, is dual to Ramsification. Compare Boolos (1985: 84–86) on supervalidity.

[4] I am glossing over some complications here. In particular, I am assuming that (when a primitive predicate for L 'logical truth' is available in the metalanguage), the schema, $L\ulcorner\phi\urcorner \to \phi$, will be available. But the only plausible counterexamples to the T-OUT schema $T\ulcorner\phi\urcorner \to \phi$ (where T is a primitive truth predicate) will occur when ϕ itself contains semantic vocabulary. (For instance, standard truth theories over PA admit $T\ulcorner\phi\urcorner \leftrightarrow \phi$ where ϕ is an arithmetical sentence.) So, when ϕ has no semantic vocabulary, we should similarly expect $L\ulcorner\phi\urcorner \to \phi$ to hold. And, by their very nature, disquotation principles for logical truth should be closed under uniform substitution of nonlogical vocabulary, so given some putative logical truth ϕ containing semantic vocabulary, there should be some ϕ° not containing semantic vocabulary such that $\phi \leftrightarrow \phi^\circ$. These issues deserve fuller treatment than is possible here, but the difficulties involved in fully formalizing Tarski's thesis and related principles against the background of an axiomatic truth theory do not affect our main point.

and logical consequence on the intended interpretation suffices for our purposes.

For the *first-order case*, taking $\mathscr{M}$ to be the class of all models and $\models$ to be the standard truth-in-a-model relation, Kreisel offers a well-known informal 'proof' of (2.1.3) by way of what has come to be known as the 'squeezing argument'. As he notes, "one reasons in mathematical practice, using the notion of consequence or of logical consequence, freely and surely" (Kreisel 1967: 153), and in so doing one implicitly accepts that if $\vdash \phi$, then ϕ is informally logically true. Furthermore, "the moment one takes it for granted that logic *applies* to structures" (1967: 154), one commits oneself to the implication from the informal logical truth of ϕ to $\models \phi$. But then, since the completeness of first-order logic gives us the implication from $\models \phi$ to $\vdash \phi$, we have the converse direction, yielding the result that ϕ is informally logically true if and only if $\models \phi$. (Or, more evocatively: 'informally rigorous' arguments about the notion of logical truth *on the intended interpretation* establish that $\{\phi \in \mathscr{L}^1 : \vdash \phi\} \subseteq \{\phi \in \mathscr{L}^1 : \phi \text{ is informally logically true}\} \subseteq \{\phi \in \mathcal{L}^1_\sigma : \models \phi\}$, and the completeness theorem demonstrating the co-extensiveness of the two formal notions tells us that the inclusions are improper; thus, $\{\phi \in \mathscr{L}^1 : \phi \text{ is informally logically true}\}$ is 'squeezed' between $\{\phi \in \mathscr{L}^1 : \vdash \phi\}$ and $\{\phi \in \mathcal{L}^1_\sigma : \models \phi\}$, pinning down its precise extension.[5])

But this argument only works in the first-order case because of the availability of a completeness theorem. The higher-order versions of (2.1.3) and (2.1.4), for appropriate model classes, will have to be shown by different means, if at all. As Kreisel notes, "*For higher order formulae we do not have a convincing proof* of e.g. [the second-order case] [...] though one would expect one" (1967: 157). The second half of this claim, however, suggests that there are some reasons, not amounting to proof, for believing that appropriate model classes exist making (2.1.3), at least, true; Kreisel fleshes out the suggestion in the textbook on mathematical logic he wrote with Jean-Louis Krivine:

> To see that this state of affairs [the absence of an informally rigorous proof] does not by itself cast doubt on the significance of Val^2, compare, e.g., our

[5]Compare the similar discussion in Kreisel (1965: 116). On squeezing arguments more generally, see Smith (2013: 348–66).

> present evidence for $Val^2 = \bar{V}^2$ with that for $Val^1 = \bar{V}^1$ before Gödel's proof. First, in both cases, whenever in actual practice one recognizes that a formula belongs to $\bar{V}^1 \cup \bar{V}^2$, one also recognizes that it is intuitively valid; and conversely. Second, for many $A \in \mathcal{L}^2$ we do not know whether $A \in \bar{V}^2$ or $A \notin \bar{V}^2$; but no more can we effectively decide for an arbitrary $A \in \mathcal{L}$, even after the proof of [the completeness of first-order logic] whether $A \in \bar{V}^1$ or $A \notin \bar{V}^1$ (or, equivalently, whether $A \in Val^1$ or $A \notin Val^1$). (Kreisel and Krivine 1967: 191)

Here $\bar{V}$ corresponds to the V of Kreisel (1967). Stewart Shapiro terms the second-order version of (2.1.3), with $\mathscr{M}$ the class of (extensional) standard second-order models, 'Kreisel's Principle'; he notes that it constitutes a "*presupposition* of the use of the set-theoretic hierarchy to provide the semantics of [higher-order] formal languages" (1987: 309). As he notes, working in ZFC^2 (and given some modest background assumptions, including Tarski's thesis), the principle is equivalent to the reflection schema $\phi \to V_\alpha \vDash \phi$, for ϕ a sentence of the language of second-order set theory, which entails the existence of n inaccessible cardinals for every $n \in \mathbb{N}$.[6] The minimal standard model of ZFC^2+ this reflection principle is, up to isomorphism, V_κ for κ the least Π^2_0-indescribable-without-parameters inaccessible cardinal (Shapiro 1987: 313, 332 n. 8).[7]

Kreisel's explanation suggests that, in the absence of an (even informally) rigorous argument, our evidence for the second-order version of (2.1.3) is primarily inductive: thus far we have failed to find a clear counterexample. Evidence of this sort is, of course, far from negligible: a significant part of our confidence in $\mathrm{Con}(\mathsf{ZFC})$, for example, is based on the fact that a century's worth of detailed investigation has failed to unearth an inconsistency. In the case of $\mathrm{Con}(\mathsf{ZFC})$, however, we also have an additional source of evidence. We take ourselves to be able to grasp, at least to an imperfect

[6] The schema was first studied by Azriel Lévy (1960a, 1960b).

[7] Considerable care is required here: indexing the αth inaccessible by i_α, it is true that, in every model, there exist i_α for all $\alpha < \kappa$ (with κ specified as above), but it does not follow from this that, if ZFC^2+ the reflection principle holds, then the universe contains i_α for $\omega \leq \alpha < \kappa$, unless we assume (2.1.3) itself in the metatheory and thereby lift results about models to results about the intended interpretation.

degree, the intended model of pure set theory—V, the iterative hierarchy.

Could we acquire a similar grasp of the intended interpretation of higher-order logic in a way that would allow us to fix on a model class and consequence relation for which we can reasonably believe that (2.1.3–4) hold? The answer to this question is unclear, but there are some reasons for optimism. As noted, under the right conditions, the version of (2.1.3) that Kreisel considers is equivalent to a weak reflection principle. More generally, any version of (2.1.3–4), when formulated over a suitable set theory in a language appropriately extending the $\mathscr{L}$ under consideration. But the basic intuition behind reflection—that the universe is large and homogeneous in such a way that, when we restrict ourselves to 'ordinary' logical resources, we encounter expressive limitations in pinning its extent down—is fundamental to modern set theory.[8] A full defence of reflection principles would lead far away from the topic at hand, but at least tentative acceptance of them is a natural complement to the generally deferential attitude to set-theoretical orthodoxy I defended in §1.1. But once we have some assurance that there exist *some* set-theoretic $\mathscr{M}$ and $\models$ for which (2.1.3–4) hold, we can then make progress—at least in some cases—by *ruling out* some candidate model classes on the basis of direct intuition about the structure of higher-order reality. For instance, if we reject predicativism or other linguistically motivated constraints on what counts as a higher-order entity, then we should reject any model class that admits models of higher-order languages for which it is not the case that, for any entities whatsoever of nth order, there is some $n+1$th-order entity that applies to them and them alone: we should reject as a candidate for faithfulness any model class that admits nonstandard Henkin models. This reflects our particular understanding of the demands reality places on the higher-order universe; a model can be a plausible surrogate only if it mirrors it in miniature. By ruling out more and more candidate model classes, we can gain additional knowledge about higher-order logic.

[8] Extraordinary resources—such as a quantifier $\exists^U$ which is *forced* to range over absolutely everything, whose reinterpretation is prohibited—will of course lead to violations of (2.1.3) and (2.1.4); on this quantifier, see Williamson (2000: 326–33). On reflection principles generally, see Koellner (2003, 2009), Wang (1977), Reinhardt (1974), and Maddy (1988a: 501–8, 1988b: 748–56).

It is often more useful to think of the process of ruling out inadmissible model classes in terms of isolating candidate *weakly faithful* model classes, in which every model satisfies the right-to-left direction of the conditional in (2.1.3–4). Let us say that a model class $\mathscr{M}$ is *weakly faithful to logical truth* (relative to a specified truth-in-a-model relation $\models$) just in case, for all $\phi \in \mathsf{Sent}(\mathscr{L})$, ϕ is logically true on the intended interpretation *only if* $\mathfrak{M} \models \phi$ for every $\mathfrak{M} \in \mathscr{M}$; likewise, $\mathscr{M}$ is *weakly faithful to logical consequence* (relative to $\models$) just in case, for all $\phi \in \mathsf{Sent}(\mathscr{L})$ and $\Gamma \subseteq \mathsf{Sent}(\mathscr{L})$, ϕ is a logical consequence of Γ on the intended interpretation *only if*, for every $\mathfrak{M} \in \mathscr{M}$, if $\mathfrak{M} \models \psi$ for all $\psi \in \Gamma$, then $\mathfrak{M} \models \phi$.

It is tempting to follow Kreisel, at least in part, and claim that the classes of standard models he considers are at least weakly faithful to logical truth for higher-order logic. Unfortunately, this is not plausible, since Kreisel considers only extensional standard models under the usual truth-in-a-model relation. This is, of course, sufficient for mathematical purposes; but if we wish to carry out general metaphysics, then a model class that excludes non-extensional models will be acceptable only if higher-order logic *on its intended interpretation*, with absolutely unrestricted subject matter, validates principles of extensionality. And it seems clear that it does not: whether we consider higher-order identity to be primitive or define it by Leibniz's Law, the claim '$\forall x(x \text{ is a cordate} \leftrightarrow x \text{ is a renate}) \rightarrow (\lambda x.x \text{ is a cordate}) = (\lambda x.x \text{ is a renate})$' should not come out true. Nor will mere *intensional equivalence* (in the sense of necessary coextensiveness) suffice for higher-order identity. We shall discuss cases involving grounding and essence operators (regimenting, roughly, '… in virtue of …' and 'it is true in virtue of the essence of … that …') in Chapter V, but nonidentity facts provide simpler examples: intuitively $(\lambda x\,.\,x = 1 \wedge x = 7) \neq (\lambda x\,.\,x = 2 \wedge x = 4)$, even though the Concepts $(\lambda x\,.\,x = 1 \wedge x = 7)$ and $(\lambda x\,.\,x = 2 \wedge x = 4)$ both have necessarily empty extension. In §2.3, I discuss how to develop a non-extensional model theory to deal with this problem, and I argue that a plausible model class can be specified that is at least weakly faithful to logical truth and logical consequence.

2.2 Type Structures and Typed Languages

Before doing so, however, it is necessary to clarify precisely what sort of higher-order language we shall work in. Since we need to move beyond second order, it is most natural to conceive of every higher-order logic of finite order as a fragment of a general *type theory*.

Type-theoretical systems take two main forms: *Church-style* (or *static*) and *Curry-style* (or *dynamic*). In Church-style typing, every term is associated with a fixed type or collection of types; in Curry-style typing, types are assigned to terms *locally*, in the context of particular proofs, via type-assignment formulas, which are themselves syntactic objects. Type-assignment formulas must obey a formal type-assignment system, with axioms and rules of inference; we can thus speak of type assignments that follow from certain assumptions (i.e., certain type-assignment formulas taken as hypotheses).[9] Although Curry-style typing is of great practical importance in certain fields, it is not well-suited to foundational inquiry in metaphysics: in a perspicuous language developed for those purposes, types should mirror natural distinctions in reality, which (barring wholesale antirealism about metaphysical structure) rules out a system in which types are local to particular proofs.[10] I shall thus consider only Church-style systems. I shall also not consider *ramified* type systems; the mathematical limitations of such systems are well-known, and they have only two natural motivations: predicativist restrictions motivated by the vicious circle principle or something analogous and hierarchical treatments of certain paradoxes (in particular Prior-style intensional paradoxes).[11]

A *type structure* $\mathscr{T}$ is a collection of type symbols (which we take to be syntactic objects), comprising some or all of those produced by recursive application of some type symbol-generating operations (making up the class $\mathfrak{O}$) to some *basic type symbols* (making up the class $\mathfrak{T}_b$). We make the usual assumptions that everything in $\mathfrak{T}_b$ is a

[9] See Hindley and Seldin (2008: 119–79) for an introduction to Curry-style typing.

[10] On applications of Curry-style typing in computer science, see Pierce (2002).

[11] See Copi (1971), Myhill (1974, 1979), Hodes (2015), and Kamareddine, Laan, and Nederpelt (2004: 19–75) for treatments of ramification; Hodes (2015) is specifically motivated by intensional paradoxes.

syntactic atom and distinct from any symbol that can be introduced by any operation in $\mathfrak{O}$. We term the pair $\mathfrak{T} = \langle \mathfrak{T}_b, \mathfrak{O} \rangle$ a *type framework*, and we term $\mathscr{T}$ a type structure generated by $\mathfrak{T}$. (Because the operations in $\mathfrak{O}$ may act upon individual members of $\mathscr{T}$, sequences of members of $\mathscr{T}$, sets of members of $\mathscr{T}$, or more esoteric constructions, we leave the precise nature of the recursive definition open for the moment.) We term the members of $\mathscr{T} - \mathfrak{T}_b$ *complex type symbols*. In general, we use τ and variants as metasyntactic variables for type symbols.

Given a type structure $\mathscr{T}$ and a class F, a *typed collection* $F_{\mathscr{T}}$ over $\mathscr{T}$ is a pair $\{F, \mathsf{type}\}$, where $\mathsf{type} : F \to \mathscr{T}$ is a *typing function* for the elements of F. We write F_τ for $\{x \in F : \mathsf{type}(x) = \tau\}$. Given two typed collections $F_{\mathscr{T}}$ and $G_{\mathscr{T}}$ over the same type structure $\mathscr{T}$, a *typed function* is a function $f_{\mathscr{T}} : F \to G$ with the property that if $x \in F_\tau$ then $f(x) \in G_\tau$. We write f_τ for $f \upharpoonright F_\tau$ and ${}^{F_{\mathscr{T}}}G_{\mathscr{T}}$ for the collection of typed functions from F to G. We shall often abuse notation and write F for $F_{\mathscr{T}}$ or f for f_τ when no confusion will result.[12]

A language $\mathscr{L}$ is said to be *typed in accordance with* $\mathscr{T}$ (or simply *over* $\mathscr{T}$) when the class of its terms, $\mathsf{Term}(\mathscr{L})$, is associated with a typed collection $\mathsf{Term}_{\mathscr{T}}(\mathscr{L})$ over $\mathscr{T}$. In a context where $\mathscr{L}$ is fixed, we shall speak of $\mathsf{Term}_\tau(\mathscr{L})$ as the *syntactic type* τ; we also use the phrase more loosely when describing characteristics shared by the members of $\mathsf{Term}_\tau(\mathscr{L})$ and those of $\mathsf{Term}_\tau(\mathscr{L}')$ for all suitably related languages $\mathscr{L}'$. In contrast, we say of the entities that are available to be values of members of $\mathsf{Term}_\tau(\mathscr{L})$ that they are *of ontological type* $\overline{\tau}$. We write $\overline{\mathscr{D}}_\tau$ for the class of entities of ontological type $\overline{\tau}$ on the intended interpretation. Where there is no risk of confusion, we use 'type' indiscriminately for 'type symbol', 'syntactic type', and 'ontological type'.

The choice of type structure to work in is of significant philosophical importance; as such, we shall devote some space to defending ours. Applied type-theoretical frameworks, particularly in computer science, make use of an extensive variety of fine-

[12] I follow Dorr (2016: 85) for terminology and definitions, who is in turn inspired in part by Benzmüller, Brown, and Kohlhase (2004: 1035) with the exception that, in virtue of using a function type rather than a relation, I admit only *nonoverlapping* typed collections (*i.e.*, I do not permit a term to have more than one type). Overlapping typed collections have their main application in Curry-style theories and can thus be ignored here.

grained options for $\mathfrak{T}_B$ and $\mathfrak{O}$. In foundational studies, however, the choices are more constrained. We require a basic syntactic type e for proper names and first-order variables. Since we intend $\overline{\mathscr{D}}_e$ to encompass absolutely every object, we can admit no other basic 'objectual' types: we cannot, for instance, have a basic type for natural numbers, as is often convenient in other contexts. Likewise, we need a basic type t for sentences. Unlike Frege, we shall not assume that $\overline{\mathscr{D}}_t$ contains only two entities. As we shall discuss below, intensional considerations motivate a far finer-grained treatment.

Nothing else is required in order to enable us to start treating higher-order types derivatively, in a way that corresponds straightforwardly to the categorial grammar of the natural-language expressions to which we appealed in §2.1 in order to motivate higher-order quantification. For instance, the association of 'Alfredo loves Violetta' with type t and 'Alfredo' and 'Violetta' with type e motivates an account on which '… loves …' (and every higher-order variable for which '… loves …' can be an instantial term) must be of some complex type $\pi(e,e,t)$, produced from the basic types e and t by some type-generating operation $\pi \in O$. But this constraint does not itself settle the nature of π; to do this, we need to explore more the space of formal options—and their metaphysical implications—more deeply.

The basic choice point is between *functional* and *relational* type theories, using type-generating operations of the forms

- π^n_F, $n \geq 1$: from types $\tau_1, \ldots, \tau_n, \tau_o$, generates the n-ary function type $(\tau_1, \ldots, \tau_n \to \tau_o)$,

and

- π^n_R, $n \geq 0$: from types $\tau_1, \ldots, \tau_n$, generates the n-ary relation type $\langle \tau_1, \ldots, \tau_n \rangle$,

respectively.

There are, however, a few other type-generating operations that have appeared in the literature (or are natural extensions of ones that have) which deserve brief comment. We could introduce the operation π^n_{PF}, $n \geq 1$, which generates from types

$\tau_1, \ldots, \tau_n, \tau_0$ the n-ary partial function type $(\tau_1, \ldots, \tau_n \rightsquigarrow \tau_0)$, or the operation π^n_T, $n \geq 2$, which generates from types $\tau_1, \ldots, \tau_n$ the n-tuple type $(\tau_1 \otimes \cdots \otimes \tau_n)$. Use of the π^n_T operations is natural in some areas of computer science, but it is difficult to find a plausible metaphysical picture that would underlie primitive types for tuples. Partial functions, on the other hand, are in some respects very natural, despite Frege's insistance on totality, and this may afford them a *prima facie* claim to consideration as part of a *comprehensive*, pellucid type theory. But for our purposes, we need not settle this question. We are seeking not to give a complete taxonomy of all the higher-order entities that there are, but to give a metaphysically perspicuous framework for dealing with the ones that are needed to characterize the nature of mathematicalia: our inquiry is fully general, in that its results will—if correct—be absolutely exceptionless, but it nonetheless has a *restricted subject matter*. In such an inquiry, we need not claim that the typed framework we work in instantiates a syntactic type for *every ontological type*. (Indeed, if there are more than denumerably many ontological types, it is hard to see how we could describe such a framework adequately.) Instead, we can carry out investigations over a limited class of types, with absolutely unrestricted quantifiers at each such type, and still achieve the generality required for serious metaphysics, as long as there is an appropriate ontological type corresponding to every syntactic type we use.

The same considerations apply to some other type-generating operations. Orthodox transfinite type theory makes use of *cumulative types*, generated by an operation π_U producing, from a set Ξ of types, the cumulative type $\bigsqcup \Xi$. For instance, the approach taken by Linnebo and Rayo (2012), when reformulated in our notation, uses $\tau_0 = t, \tau_1 = e, \tau_{\alpha+1} = \pi^1_R(\tau_n)$, and $\hat{\tau}_\alpha = \bigsqcup\{\tau_\beta : \beta < \alpha\}$ for all α, and sets the type structure $\mathscr{T}_{Tr} = \{\tau_0, \tau_1\} \cup \{\tau_\alpha : \alpha \in \mathsf{On} \wedge \alpha \geq 2\}$. As they note, this framework permits a native (higher-order) semantics for the entire finite fragment at level $\omega + 2$; more generally, given reasonable background assumptions, a truth definition for a language over the fragment of $\mathscr{T}^C_{Tr}$ up to level α can be given at level $\alpha + 1$ if α is a successor ordinal and $\alpha + 2$ if α is a limit ordinal (Linnebo and Rayo 2012: 299–300). But we shall never need to do explicit semantics in this way for more than a fragment of the type hierar-

chy at once; thus a system incorporating only unboundedly high finite rank, with no cumulation, will suffice for our purposes.[13] There are natural extensions of π^n_R and π^n_F to π^α_R and π^α_F, with α possibly transfinite. These are appealing, and can do some of the semantic work of cumulative transfinite types in a more elegant way,[14] but are again unnecessary for our purposes. Alonzo Church's (1951a) logic of sense and denotation (LSD) incorporates a type-forming operation π_{ASC} which, from a type τ, generates the ascendant type τ'. Here the basic picture is that if a term **t** has as its semantic value an entity of type τ, then its sense is an entity of type τ'.[15] Plausibly, some version of LSD remains the most promising approach for treating the hyperintensional behaviour of terms in belief contexts (Church 1951b, Kaplan 1968: 185), but LSD is not required for the kind of hyperintensionality that is relevant to *metaphysical* work: senses are not fundamental, worldly entities.[16] Senses are presumably too fine-grained to function as desired with the hyperintensional operators of interest to metaphysics, such as grounding and essence.[17] In §3.2, I shall have occasion to appeal to a Fregean picture of sense and semantic value in arguing against neologicist theories of content, but this is done entirely informally (and the senses involved are sufficiently few in number that no problems arise from assuming that they are objects). Finally, we could admit a type-forming operation π_{PL} which, from a type τ, generates the plural type $\tau\tau$. But, for the reasons discussed in §2.1, even if we were to admit π_{PL}, we should still need other type-generating operations, since higher-order quantification is not reducible to plural quantification. As is the case with senses, we can argue informally using plurals with-

[13] It should be noted that cumulativity also conflicts with what is often thought to be a fundamental intuition behind the primitivist approach to higher-order logic: the world is rigidly type-segregated at the deepest level, with every entity occupying only one type. A version of this intuition, for instance, motivates Frege's original characterization of functions as 'unsaturated': if an object could *also* be a function, the Fregean distinction would break down.

[14] I owe this point to Jeremy Goodman.

[15] See generally Church (1951a, 1973, 1974, 1993), Anderson (1977, 1980, 1984, 1987, 1998), Kaplan (1964, 1975), and Klement (2010) for more on LSD. Note that, if we take $\overline{\mathscr{D}}_{\tau_1}$ to be disjoint from $\overline{\mathscr{D}}_{\tau_2}$ for $\tau_1 \neq \tau_2$, we must reject what Walsh (2016: 287) calls the *senses are objects axiom*, unlike most treatments of the sense/denotation distinction. For the cardinality problems with the senses are objects axioms, see, in addition to Walsh (2016), Klement (2002, 2003).

[16] But see Diehl (MS), who argues that senses can be genuinely worldly.

[17] Reasons to think that senses are extremely fined-grained derive from cases such as Mates's puzzle (Mates 1952) and Kripke's Pierre problem (Kripke 1979).

out incorporating them in our official system (and, in fact, we have done so throughout this chapter and the last); if required, we can always simulate plural quantification of type $\tau\tau$ with quantification over an appropriately rigidified entity of some higher-type $\pi(\tau)$.

We thus return to the central choice between functional and relational type theory. The most common methods of developing type hierarchies with these operations are limited in ways that render them problematic for our purposes. The canonical presentation of functional type theory is due to Church (1940): this hierarchy, which we shall term $\mathscr{T}^{Ch}$, is constructed with $\mathfrak{T}_b = \{e, t\}$ and $\mathfrak{O} = \{\pi^1_F\}$. Relations are equated with functions of the type $\overline{(\tau \to t)}$ for some τ, and n-ary functions are reduced by schönfinkelization: thus a regimentation of '… loves …' is treated as being of syntactic type $(e \to (e \to t))$. In contrast, the standard formulation of relational type theory is due to Stephen Orey (1959): $\mathscr{T}^{Or}$ is constructed with $\mathfrak{T}_b = \{e\}$ and $\mathfrak{O} = \{\pi^n_R : n \in \mathbb{N}\}$: the type resulting form the action of π^0_R on no inputs, $\langle\rangle$, is in effect treated as the syntactic analogue of the ontological type $\bar{t}$ (the type of propositions). In particular, expressions of type $\langle\rangle$ without free variables are taken to be evaluable as true or false. On an Orey-style treatment, functions are identified with corresponding relations: thus, for example, what would be quantification over type-$\overline{(\tau \to \tau)}$ entities in a Church-style framework is replaced by quantification over type-$\overline{\langle\tau, \tau\rangle}$ entities with added totality and right uniqueness restrictions.

The central limitation of $\mathscr{T}^{Ch}$ lies in the use of schönfinkelization. It is natural to motivate type theory as an extension of conventional second-order logic, with quantification into predicate and sentence position, but the fragment of $\mathscr{T}^{Ch}$ admitting quantification over types of 'at most second order' (i.e., e and $(e \to t)$) corresponds only to *monadic* second-order logic, which is strictly weaker than full second-order logic (Shapiro 1991: 220–26). It is possible, given the existence of infinitely many objects and a background choice principle, to simulate quantification over relations of adicity above two in the fragment of $\mathscr{T}^{Ch}$ that admits quantification over types $e, (e \to t)$, and $((e \to t) \to t)$ (Burgess, Hazen, and Lewis 1991; Hazen 1997a, 1997b), but the pro-

cedure is cumbrous. Moreover, although the assumption of a choice principle in the background theory is unproblematic from the point of view defended in this work—the point of view, that is, of the mathematical realist who thinks that choice principles of arbitrarily high type are true on their intended interpretation—the other requirement is more problematic. Even though there are infinitely many objects, there will be occasions when implicit quantifier restrictions are in play and the assumption of an infinite domain is inappropriate: in those cases there is no bound to the complexity of the types in $\mathscr{T}^{Ch}$ needed to represent full second-order logic. These technical considerations give support to the intuitive appeal of Frege's claim that "instead of functions of two arguments we can deal with functions of a single but complex argument; but the distinction between functions of one and two arguments still holds in all its sharpness" (Frege 1891: *31/1960: 41)—in other words, that schönfinkelization is merely a useful trick that occults deep metaphysical structure.[18]

For these reasons, $\mathscr{T}^{Or}$ is in many ways a more elegant and natural framework. It does, however, have two drawbacks. First, the failure to implement function types directly means that $\mathscr{T}^{Or}$ can serve as a framework for languages formalizing mathematics only at the price of substantial violence to surface form: functions from objects to objects, functions from functions to objects, and so on appear extensively in both ordinary mathematical practice and the kind of foundational explorations at issue here. In Chapter IV, for instance, I defend an account of ordinals and cardinals on which they correspond to the values of primitive abstraction functors. But, without functional types, there is no way to implement an abstraction functor directly: instead, abstraction principles must be expressed involving predicates for relations with explicit totality and right uniqueness requirements.

Secondly, the treatment of propositions in $\mathscr{T}^{Or}$ as non-basic is perplexing: it is far from obvious why the ontological type of propositions should be related to that of objects by the action of a nullary type-generating operation. At least, there seems to be

[18] "Man kann auch statt der Funktionen mit zwei Argumenten Funktionen eines einzigen, aber komplexen Arguments betrachten, wobei jedoch der Unterschied zwischen den Funktionen mit einem und denen mit zwei Argumenten in ganzer Schärfe bestehen bleibt" (trans. Geach).

nothing in the basic analogy that we appeal to in order to explain ontological types by way of their syntactical correlates—the thesis that a sentence stands to a proposition as a proper name stands to an object—that gives us reason to expect that this would be the case. In contrast, the characterization of relations that underlies theories with $\mathfrak{T}^B = \{e, t\}$ and $\mathfrak{O} \subseteq \{\pi^1_F : n \geq 1\}$ renders priority in the order of nature consonant with priority in the order of the understanding: the basic types are the two to which we appear to have the most direct access, and derived types result from the application of a familiar procedure, the concept of which is a simple generalization of the notion of function already familiar from ordinary mathematics.[19] These considerations give us reason to believe in the existence of functional types. But they do not answer the question whether we ought to accept *sui generis* primitive relational types as well, alongside types for propositional functions. Here an answer similar to that given for partial functional and plural types is appropriate: since our purpose is to develop a framework suitable for investigating the nature of mathematical objects, not to give a comprehensive account of all the ontological types that there are, we can accept the reduction of relations to propositional functions at least for practical purposes.

We can thus finally choose a type hierarchy: we define $\mathscr{T}$ as the full type hierarchy based on $\mathfrak{T}_b = \{e, t\}$ and $\mathfrak{O} = \{\pi^n_F : n \geq 1\}$. In other words, $e, t \in \mathscr{T}$; if $\tau_1, \ldots, \tau_n, \tau_o \in \mathscr{T}$, then $(\tau_1, \ldots, \tau_n \to \tau_o) \in \mathscr{T}$; and nothing else is in $\mathscr{T}$. We can then define the set $\mathscr{T}^R \subsetneq \mathscr{T}$ of *relational types* (understood, whether by proxy or not, as propositional functional types) as follows: $e, t \in \mathscr{T}^R$; if $\tau_1, \ldots, \tau_n \in \mathscr{T}^R$, then $(\tau_1, \ldots, \tau_n \to t) \in \mathscr{T}^R$; and nothing

[19] There is one additional argument for $\mathscr{T}^{Or}$ over a functional type hierarchy, however, which should be addressed: Oppenheimer and Zalta (2011) have maintained that, in spite of appearances, the equation of types in $\mathscr{T}^{Or}$ with those in a $\mathscr{T}^{Ch}$ by the mapping $\langle \tau_1, \ldots, \tau_n \rangle \mapsto (\tau_1 \to \cdots \to (\tau_n \to t))$ (or those in our hierarchy $\mathscr{T}$, set out below, by the simpler mapping $\langle \tau_1, \ldots, \tau_n \rangle \mapsto (\tau_1, \ldots, \tau_n \to t)$) is inadequate even for practical purposes. Their argument depends on the claim that we can make sense of *two* kinds of atomic predication, as used in the 'object theory' of Zalta (1983, 1988), exemplification and encoding, but that there is no corresponding distinction to be found in types of function application. According to (one version of) object theory, for example, Sherlock Holmes *encodes but does not exemplify* the property of being a detective, but *exemplifies but does not encode* the property of being an abstract object. I do not deny that there is no corresponding distinction in types of function application; I deny, however, that the purported distinction in types of atomic predication is coherent. For I do not know what it can mean to say that Sherlock Holmes encodes but does not exemplify the property of being a detective except that, although Sherlock Holmes is not a detective (for no abstract object is a detective), it is nonetheless the case that, *according to the Sherlock Holmes stories*, Sherlock Holmes is a detective. Here, each 'is' is the ordinary 'is' of predication: the work of the theory of fiction is concentrated in the 'according to … ' operator.

else is in $\mathscr{T}^R$. We define the *complexity* of a type, $\mathsf{cpx}(\tau)$ as follows: $\mathsf{cpx}(e) = \mathsf{cpx}(t) = 0$; $\mathsf{cpx}((\tau_1,\ldots,\tau_n \to \tau_0)) = \max(\{\mathsf{cpx}(\tau_i) : 0 \leq i \leq n\}) + 1$. (From now on, we use $\mathscr{T}$ only as the symbol for this particular type hierarchy, not as a schematic letter for type hierarchies in general.) The class of complex types, $\mathscr{T}^{Cpx}$, is $\mathscr{T} - \mathfrak{T}_b$. In Chapter VI, when we discuss higher-order types at length, we shall have occasion to introduce some further abbreviations for type symbols to reduce space, but they will not be necessary here. In general, if $\mathbf{t}$ is a term of type τ, we write $\mathbf{t}^\tau$ with the type index as a superscript on its initial occurrence in an expression to indicate that $\mathsf{type}(\mathbf{t}) = \tau$. In subsequent occurences in the same expression, $\mathbf{t}$ is presented bare to avoid notational clutter.

An alphabet $\mathfrak{A}$ is the carrier set for a typed collection $\mathfrak{A}_{\mathscr{T}}$, comprising: (1) for each type $\tau \in \mathscr{T}$, denumerably many variables $x_0^\tau, x_1^\tau, \ldots$; (2) for each type $\tau \in \mathscr{T}$, *at most* denumerably many logical constants $l_0^\tau, l_1^\tau, \ldots$; (3) for each type $\tau \in \mathscr{T}$, *at most* denumerably many nonlogical constants $c_0^\tau, c_1^\tau, \ldots$. We define $\mathsf{Var}_\tau = \{x_0^\tau, x_1^\tau, \ldots\}$; $\mathsf{LCons}_\tau = \{l_0^\tau, l_1^\tau, \ldots\}$; $\mathsf{Cons}_\tau = \{c_0^\tau, c_1^\tau, \ldots\}$; $\mathsf{Var} = \bigcup_{\tau\in\mathscr{T}} \mathsf{Var}_\tau$; $\Sigma_L = \bigcup_{\tau\in\mathscr{T}} \mathsf{LCons}_\tau$; $\Sigma_C = \bigcup_{\tau\in\mathscr{T}} \mathsf{Cons}_\tau$; $\Sigma = \Sigma_C \cup \Sigma_L$.

In general, we use the following conventions. Normally, we let x, y, z, and decorated variants replace x_i^τ for various τ subject to the rule about subsequent type indications in an expression given above. Where no type subscript is indicated on x, y, z, or a decorated variant *anywhere in the expression in which it figures*, however, the symbol always stands for a variable of type e. We write X^n, Y^n, Z^n, and decorated variants for $x_0^{(e,\cdots,e\to t)}, x_1^{(e,\cdots,e\to t)}, \ldots$, where $(e,\cdots,e\to t)$ abbreviates $(\underbrace{e,\cdots,e}_{n \text{ times}} \to t)$. We write X, Y, Z, and so on, with no superscript, for X^1, Y^1, Z^1, and variants. We write p, q, and variants for x_0^t, x_1^t. We sometimes use f^n and decorated variants for $x_0^{(e,\cdots,e\to e)}, x_1^{(e,\cdots,e\to e)}, \ldots$, where $(e,\cdots,e\to e)$ abbreviates $(\underbrace{e,\cdots,e}_{n \text{ times}} \to e)$. We allow various ad hoc symbols to stand for members of Σ; the standard symbols for logical constants, which stand for symbols in Σ_L, are a particularly important case. In general, we use boldface $\mathbf{x}, \mathbf{y}, \ldots, \mathbf{l}, \mathbf{c}, \mathbf{t}, \mathbf{p}$ and variants, with appropriate type indices, as schematic letters for variables, logical constants, nonlogical constants, terms, and formulas, respectively. (In general, $\mathbf{x}, \mathbf{y}, \ldots,$

may be of any type, not just *e*.) We also occasionally use $\phi, \psi, \ldots$, for formulas, with free variables (or some portion thereof) displayed. These rules apply only in expressions of our official language: elsewhere in prose explanations, we freely use symbols of all kinds in accordance with the conventions of ordinary mathematical English.

A language $\mathscr{L}^{\Sigma}$ is determined by Σ, the component of the alphabet $\mathfrak{A}$ that is allowed to vary; we term Σ the *signature* of $\mathscr{L}^{\Sigma}$. (This is somewhat broader than the standard use of the term, where a signature comprises only nonlogical constants.) Formation rules for $\mathscr{L}^{\Sigma}$ are as follows:

- If $\mathbf{x}^{\tau} \in \mathsf{Var}_{\tau}$, then $\mathbf{x}^{\tau} \in \mathsf{Term}_{\tau}(\{\mathbf{x}^{\tau}\})$;
- If $\mathbf{l}^{\tau} \in \mathsf{LCons}_{\tau}$, then $\mathbf{l}^{\tau} \in \mathsf{Term}_{\tau}(\emptyset)$;
- If $\mathbf{c}^{\tau} \in \mathsf{Cons}_{\tau}$, then $\mathbf{c}^{\tau} \in \mathsf{Term}_{\tau}(\emptyset)$;
- If $\mathbf{t}_{\mathrm{o}}^{(\tau_1,\ldots,\tau_n\to\tau_{\mathrm{o}})} \in \mathsf{Term}_{(\tau_1,\ldots,\tau_n\to\tau_{\mathrm{o}})}(\Gamma_{\mathrm{o}})$, $\mathbf{t}_1^{\tau_1} \in \mathsf{Term}_{\tau_1}(\Gamma_1), \ldots$, $\mathbf{t}_n^{\tau_n} \in \mathsf{Term}_{\tau_n}(\Gamma_n)$, then $\mathbf{t}_{\mathrm{o}}^{(\tau_1,\ldots,\tau_n\to\tau_{\mathrm{o}})}\mathbf{t}_1^{\tau_1}\cdots\mathbf{t}_n^{\tau_n} \in \mathsf{Term}_{\tau_{\mathrm{o}}}(\bigcup_{\mathrm{o}\leq i\leq n}\Gamma_i)$;
- If $\mathbf{x}_1^{\tau_1} \in \mathsf{Var}_{\tau_1}, \ldots, \mathbf{x}_n^{\tau_n} \in \mathsf{Var}_{\tau_n}$, $\mathbf{t}_{\mathrm{o}}^{\tau_{\mathrm{o}}} \in \mathsf{Term}_{\tau_{\mathrm{o}}}(\Gamma)$, and $\mathbf{x}_1^{\tau_1}, \ldots, \mathbf{x}_n^{\tau_n} \in \Gamma$, then $(\lambda \mathbf{x}_1^{\tau_1}\cdots\mathbf{x}_n^{\tau_n}\,.\,\mathbf{t}_{\mathrm{o}}^{\tau_{\mathrm{o}}}) \in \mathsf{Term}_{(\tau_1,\ldots,\tau_n\to\tau_{\mathrm{o}})}(\Gamma - \{\mathbf{x}_1^{\tau_1}, \ldots, \mathbf{x}_n^{\tau_n}\})$.

We define $\mathsf{Term}_{\tau} = \bigcup\{\mathsf{Term}_{\tau}(\Gamma) : \Gamma \subseteq \mathsf{Var}\}$ and $\mathsf{Term} = \bigcup_{\tau\in\mathscr{T}}\mathsf{Term}_{\tau}$. If $\mathbf{t} \in \mathsf{Term}_{\tau}(\{\mathbf{x}_1^{\tau_1}, \ldots, \mathbf{x}_n^{\tau_n}\})$, then we say $\mathbf{t}$ is a *term of type* τ *with* $\mathbf{x}_1^{\tau_1}, \ldots, \mathbf{x}_n^{\tau_n}$ *free.* A term of type *t* is a *formula.* A formula with no free variables is a sentence.

In two respects other than the acceptance of polyadic function terms, this presentation of the syntax of functional type theory differs from more standard ones. First, we do not allow vacuous λ-binding; for a term of the form $(\lambda \mathbf{x}_1^{\tau_1}\cdots\mathbf{x}_n^{\tau_n}\,.\,\mathbf{t}_{\mathrm{o}}^{\tau_{\mathrm{o}}})$ to be well-formed, $\mathbf{x}_1^{\tau_1}, \ldots, \mathbf{x}_n^{\tau_n}$ must occur free in $\mathbf{t}_{\mathrm{o}}^{\tau_{\mathrm{o}}}$. Languages with this property are sometimes termed $\lambda\mathbf{I}$-languages, as opposed to $\lambda\mathbf{K}$-languages, which do allow vacuous λ-binding. As Dorr (2016: 57) notes, the utility of vacuous λ-binding is largely dependent on acceptance of vacuous β-conversion, which imposes a strong degree of extensionality on the system.[20] Since we do not wish to build this constraint into the type structure, we

[20] See Dorr (2016: 82) for some consequences of vacuous β-conversion.

follow his suggestion that "[t]he opponents of vacuous β-conversion may legitimately take $\lambda\mathbf{I}$-languages to be metaphysically more perspicuous than $\lambda\mathbf{K}$-languages, as well as formally more convenient" (2016: 58).

Secondly, we do not specify a fixed choice of logical constants in advance. The reason for this is that, once hyperintensional operators such as grounding and essence (or even higher-order identity) are in play, we can no longer assume that logically equivalent expressions can be substituted everywhere *salva veritate*. This fact will be important in §5.3, in connexion with the distinction between logically equivalent primitives and composite expressions in grounding contexts. We thus want to allow for the possibility that a language may contain a logical constant corresponding to any genuinely logical operator, in the standard sense of permutation-invariance (Tarski 1986, Sher 1991, McGee 1996). We implement this through the notion of a *logic-correct* model, defined below.

First, however, we need to set out a few technical notions concerning the syntax of $\mathscr{L}^{\Sigma}$ that will be required in §2.3.

Definition 2.2.1. We define *uniform simultaneous capture-free substitution* on terms of a language $\mathscr{L}^{\Sigma}$ standardly; the details can be found in Dorr (2016: 90) or Hindley and Seldin (2008: 7). We write $\mathbf{t}_0[\mathbf{t}_1/\mathbf{t}'_1]$ and $\mathbf{t}_0[\mathbf{t}_1,\ldots,\mathbf{t}_n/\mathbf{t}'_1,\ldots,\mathbf{t}'_n]$ for the result of performing uniform simultaneous capture-free substitution of $\mathbf{t}_1$ for $\mathbf{t}'_1$ (resp. $\mathbf{t}_1,\ldots,\mathbf{t}_n$ for $\mathbf{t}'_1,\ldots,\mathbf{t}'_n$) (when $\mathbf{t}'_1,\ldots,\mathbf{t}'_n \in \Sigma \cup \mathrm{Var}$). We say that $\mathbf{t}_1$ *immediately α-reduces* to $\mathbf{t}_2$ ($\mathbf{t}_1 \blacktriangleright_\alpha \mathbf{t}_2$) if $\mathbf{t}_1 = (\lambda\mathbf{x}_1\cdots\mathbf{x}_n \,.\, \mathbf{t}_0)$ and $\mathbf{t}_2 = (\lambda\mathbf{y}_1\cdots\mathbf{y}_n \,.\, \mathbf{t}_0[\mathbf{y}_1,\ldots,\mathbf{y}_n/\mathbf{x}_1,\ldots,\mathbf{x}_n])$ where either $\mathbf{y}_i = \mathbf{x}_i$ or $\mathbf{y}_i$ is not free in $\mathbf{t}_0$; that $\mathbf{t}_1$ *immediately β-reduces* to $\mathbf{t}_2$ ($\mathbf{t}_1 \blacktriangleright_\beta \mathbf{t}_2$) if $\mathbf{t}_1 = (\lambda\mathbf{x}_1\cdots\mathbf{x}_n \,.\, \mathbf{t}_0)\mathbf{t}'_1\cdots\mathbf{t}'_n$ and $\mathbf{t}_2 = [\mathbf{t}'_1,\ldots,\mathbf{t}'_n/\mathbf{x}_1,\ldots,\mathbf{x}_n]\mathbf{t}_0$; and that $\mathbf{t}_1$ *immediately η-reduces* to $\mathbf{t}_2$ ($\mathbf{t}_1 \blacktriangleright_\eta \mathbf{t}_2$) if $\mathbf{t}_1 = (\lambda\mathbf{x}_1\cdots\mathbf{x}_n \,.\, \mathbf{t}_0)$, $\mathbf{t}_0 = \mathbf{t}_2\mathbf{x}_1\cdots\mathbf{x}_n$, and none of $\mathbf{x}_1,\ldots,\mathbf{x}_n$ is free in $\mathbf{t}_2$. For $\star \subseteq \{\alpha,\beta,\gamma\}$, we say that $\mathbf{t}_1$ *immediately $\star$-reduces* to $\mathbf{t}_2$ if $\mathbf{t}_1$ ($\mathbf{t}_1 \blacktriangleright_\star \mathbf{t}_2$) if $\mathbf{t}_1$ *immediately $\bullet$-reduces* to $\mathbf{t}_2$ for some $\bullet \in \star$. (We write $\blacktriangleright_{\alpha\beta}$ for $\blacktriangleright_{\{\alpha,\beta\}}$ and so on, and likewise for all other symbolism in this section.) We say

that $\mathbf{t}_1$ *one-step* $\star$*-reduces* to $\mathbf{t}_2$ ($\mathbf{t}_1 \blacktriangleright_{1\star} \mathbf{t}_2$) if $\mathbf{t}_0 \blacktriangleright_\star \mathbf{t}'_0$, $\mathbf{t}_0$ is a constituent of $\mathbf{t}_1$, and $\mathbf{t}_2$ results from replacing $\mathbf{t}_0$ with $\mathbf{t}'_0$. Furthermore, $\mathbf{t}_1$ *is one-step* $\star$*-equivalent* to $\mathbf{t}_2$ ($\mathbf{t}_1 \approx_{1\star} \mathbf{t}_2$) if $\mathbf{t}_1 \blacktriangleright_{1\star} \mathbf{t}_2$ and $\mathbf{t}_2 \blacktriangleright_{1\star} \mathbf{t}_1$. Finally, $\mathbf{t}_a$ $\star$*-reduces* to $\mathbf{t}_b$ ($\mathbf{t}_a \rhd_\star \mathbf{t}_b$) if there is a finite sequence $\langle \mathbf{t}_1, \ldots, \mathbf{t}_n \rangle$ such that $\mathbf{t}_a \blacktriangleright_{1\star} \mathbf{t}_1$, $\mathbf{t}_i \blacktriangleright_{1\star} \mathbf{t}_{i+1}$ for $1 \leq i \leq n$, and $\mathbf{t}_n \blacktriangleright_{1\star} \mathbf{t}_b$; $\mathbf{t}_a$ *is* $\star$*-equivalent* to $\mathbf{t}_b$ ($\mathbf{t}_a \approx_\star \mathbf{t}_b$) if there is a finite sequence $\langle \mathbf{t}_1, \ldots, \mathbf{t}_n \rangle$ such that $\mathbf{t}_a \approx_{1\star} \mathbf{t}_1$, $\mathbf{t}_i \approx_{1\star} \mathbf{t}_{i+1}$ for $1 \leq i \leq n$, and $\mathbf{t}_n \approx_{1\star} \mathbf{t}_b$.[21]

We now define the basic notion of an *extensional frame*. Extensional frames are the basis of extensional models for higher-order languages, but they have much broader utility: every form of non-extensional model-theoretic structure we shall consider for a language $\mathscr{L}^\Sigma$ can be transformed into an extensional frame by quotienting out non-extensional structure.

Definition 2.2.2. (Extensional Frames) An *extensional frame* is a typed collection $D_\mathscr{T}$ where $D_t = \{0, 1\}$, D_e is a nonempty set such that $D_e \cap D_t = \emptyset$, and $D_{(\tau_1,\ldots,\tau_n \to \tau_0)} \subseteq {}^{D_{\tau_1} \times \cdots \times D_{\tau_n}} D_{\tau_0}$. If, for every $\tau_0, \ldots, \tau_n \in \mathscr{T}$, $D_{(\tau_1,\ldots,\tau_n \to \tau_0)} = {}^{D_{\tau_1} \times \cdots \times D_{\tau_n}} D_{\tau_0}$, then the frame is *standard*. We write $D = \bigcup_{\tau \in \mathscr{T}} D_\tau$.

We abbreviate 'standard extensional frame' as *SE-frame*. We occasionally use $\mathscr{F}$ as a variable ranging over frames, writing $D(\mathscr{F}), D_\tau(\mathscr{F})$, and the like, with the obvious meaning. For the reasons discussed in §2.1, our main interest is in standard frames.

On this basis, we can characterize logical constants in a way that allows extensive freedom as to the contents of Σ_L. So long as extensionality is in place, the standard permutation-invariance test can easily be applied to entities of higher type. Let $\mathsf{Perm}(C)$ denote the class of permutations on a set C. Given a standard extensional frame $D_\mathscr{T}$ and a permutation $\pi \in \mathsf{Perm}(D_e)$, we define the typed function $\hat{\pi} : D \to D$ generating *permutations induced by* π as follows: $\hat{\pi}_t = \mathrm{id}$; $\hat{\pi}_e = \pi_0$; $\hat{\pi}_{(\tau_1,\ldots,\tau_n \to \tau_0)}$ is the function taking the function $\langle x_1, \ldots, x_n \rangle \mapsto y$ to $\langle \hat{\pi}(x_1), \ldots, \hat{\pi}(x_n) \rangle \mapsto \hat{\pi}(y)$. Let $\mathfrak{Fr}$ denote the class of SE-frames. An *extensional candidate value (EC-value) of type* τ is a function

[21] I follow Dorr (2016: 90–91), with some modifications, for these definitions.

$f : \mathfrak{Fr} \to \bigcup\{D(\mathscr{F}) : \mathscr{F} \in \mathfrak{Fr}\}$ such that $f(\mathscr{F}) \in D_\tau(\mathscr{F})$. An EC-value f is *permutation-invariant* just in case, for every $\mathscr{F} \in \mathfrak{Fr}$, for every $\pi \in \mathsf{Perm}(D_e(\mathscr{F}))$, $\hat{\pi}(f(\mathscr{F})) = f(\mathscr{F})$. We define the class $\mathfrak{Log}_\tau$ of *logical EC-values of type* τ as the class containing every permutation-invariant EC-value of type τ and set $\mathfrak{Log} = \bigcup_{\tau \in \mathscr{T}} \mathfrak{Log}_\tau$.

Even in the extensional setting, this characterization of $\mathfrak{Log}$ may be thought unduly promiscuous by standard criteria. For instance, McGee's "wombat disjunction" (1996: 575–76)—in our terms, the candidate value f of type $(t, t \to t)$ such that (1) if $D_e(\mathscr{F})$ contains wombats, $f(\mathscr{F})(x, y) = 1$ if and only if $x = 1$ or $y = 1$ and (2) if $D_e(\mathscr{F})$ does not contain wombats, $f(\mathscr{F})(x, y) = 1$ if and only if $x = 1$ and $y = 1$—is an element of $\mathfrak{Log}$. We could avoid this problem by moving to invariance under arbitrary bijections, requiring comparisons across standard extensional models (Sher 1991: 61–65; McGee 1996: 576), but this is not necessary for our purposes. All that is required is that $\mathfrak{Log}$ contains the EC-value for all standard logical constants and any additional resources (generalized quantifiers, primitive terms corresponding to Concepts such as injectivity and the like, and so on) that we may need later. If we need some unusual primitive logical constant $\mathbf{l}$, we can always explicitly specify both an informal interpretation of $\mathbf{l}$ and an explicit characterization of the logical EC-value $\ell(\mathbf{l})$ assigned to $\mathbf{l}$, so long as it is evident that $\ell(\mathbf{l}) \in \mathfrak{Log}$, that $\ell(\mathbf{l})$ is the natural representation in an extensional framework for the intended interpretation of $\mathbf{l}$, and that, on its intended interpretation, $\mathbf{l}$ counts as a logical constant even by standards more restrictive than bare permutation-invariance.

We say that a language $\mathscr{L}^\Sigma$ is *logically protointerpreted* just in case $\mathscr{L}^\Sigma$ is associated with a typed function $\ell : \Sigma_L \to \mathfrak{Log}$. We also define a few specific logical constants that we shall expect always to be present in Σ_L for any language we are interested in. We let $\neg$ be a logical constant of type $(t \to t)$, with $\ell(\neg)(\mathscr{F})$, for $f \in \mathfrak{Fr}$, the function $f \in D_{(t \to t)}(\mathscr{F})$ such that $f(p) = 1$ if and only if $p = 0$. We let $\wedge$ and $\vee$ be logical constants of type $(t, t \to t)$, with $\ell(\wedge)(\mathscr{F})$ the function $f \in D_{(t,t \to t)}(\mathscr{F})$ such that $f(\langle x, y\rangle) = 1$ if and only if $x = y$ and $y = 1$; $\ell(\vee)(\mathscr{F})$ is the function $f \in D_{(t,t \to t)}(\mathscr{F})$ such that $f(\langle x, y\rangle) = 0$ if and only if $x = 0$ and $y = 0$. For every $\tau_1, \ldots, \tau_n \in \mathscr{T}$, we let $\Pi_{(\tau_1, \ldots, \tau_n)}$

be a logical constant of type $(((\tau_1,\ldots,\tau_n)\to t)\to t)$ with $\ell(\Pi_{(\tau_1,\ldots,\tau_n)})(\mathscr{F})$ the function $f\in D_{(((\tau_1,\ldots,\tau_n)\to t)\to t)}(\mathscr{F})$ such that, for every $g\in D_{((\tau_1,\ldots,\tau_n)\to t)}(\mathscr{F})$, $f(g)=1$ if and only if, for every $x_1\in D_{\tau_1}(\mathscr{F}),\ldots,x_n\in D_{\tau_1}(\mathscr{F})$, $g(\langle x_1,\ldots,x_n\rangle=1$. Finally, for every $\tau\in\mathscr{T}$, we let $=_\tau$ be a logical constant of type $(\tau,\tau\to t)$ with $\ell(=_\tau)(\mathscr{F})$ the function $f\in D_{(\tau,\tau\to t)}(\mathscr{F})$ such that, for all $x,y\in D_\tau(\mathscr{F})$, $f(\langle x,y\rangle=1$ if and only if $x=y$.

We introduce a few abbreviations into any language containing these constants: we write $(\mathbf{p}_1\vee\mathbf{p}_2)$, $(\mathbf{p}_1\wedge\mathbf{p}_2)$, and $\mathbf{t}_1^\tau=\mathbf{t}_2^\tau$ for $\vee\mathbf{p}_1\mathbf{p}_2$, $\wedge\mathbf{p}_1\mathbf{p}_2$, and $=_\tau\mathbf{t}_1^\tau\mathbf{t}_2^\tau$, respectively, and we write $\forall\mathbf{x}_1^{\tau_1}\cdots\forall\mathbf{x}_n^{\tau_n}\phi(\mathbf{x}_1^{\tau_1},\ldots,\mathbf{x}_n^{\tau_n})$ for $\Pi_{(\tau_1,\ldots,\tau_n)}(\lambda\mathbf{x}_1^{\tau_1}\cdots\mathbf{x}_n^{\tau_n}\,.\,\phi)$. (We could define higher-type identity using Leibniz's Law as $\mathbf{t}_1^\tau=\mathbf{t}_2^\tau\overset{\text{Df.}}{=}\forall x^{(\tau\to t)}(x\mathbf{t}_1^\tau\leftrightarrow x\mathbf{t}_2^\tau)$, but it is more natural to take it as primitive; in any case, all models we are interested in are *Leibnizian* (Dorr 2016: 100): the primitive and defined notion coincide.) We also occasionally write $\mathbf{t}_1^\tau\equiv\mathbf{t}_2^\tau$ for $\mathbf{t}_1^\tau=\mathbf{t}_2^\tau$ if $\tau\neq e$. We let $\mathsf{BasLC}=\{\neg,\vee,\wedge\}\cup\{=_\tau\colon\tau\in\mathscr{T}\}\cup\{\Pi_{(\tau_1,\ldots,\tau_n)}:\tau_1,\ldots,\tau_n\in\mathscr{T}\}$. We say that $\mathscr{L}^\Sigma$ is *logically interpreted* just in case $\mathscr{L}^\Sigma$ is logically protointerpreted, $\mathsf{BasLC}\subseteq\Sigma_L$, and for each $\star\in\mathsf{BasLC}$, $\ell(\star)$ has the value specified in this paragraph.

We now define the notion of an extensional model and set out the standard extensional type-theoretic semantics for such models.

Definition 2.2.3. (Extensional Models) An *extensional model* for a logically interpreted language $\mathscr{L}^\Sigma$ is a triple $\mathfrak{M}=\langle D_{\mathscr{T}},v_L,v_C\rangle$, where $D_{\mathscr{T}}$ is an extensional frame and $v_L:\Sigma_L\to D$ and $v_C:\Sigma_C\to D$ are typed functions. (We write $D_{\mathscr{T}}(\mathfrak{M})$ or $D_{\mathscr{T}}^{\mathfrak{M}}$, and so on, when needed for clarity, with the obvious interpretation. As usual, $D=\bigcup_{\tau\in\mathscr{T}}D_\tau$.) An extensional model $\mathfrak{M}$ is standard if $D_{\mathscr{T}}(\mathfrak{M})$ is standard. An extensional model $\mathfrak{M}$ is *logic-correct* if, for every $\mathbf{l}\in\Sigma_L$, $v_L(\mathbf{l})=\ell(D_{\mathscr{T}}((\mathfrak{M}))(\mathbf{l})$. We denote the class of all standard, logic-correct extensional models for $\mathscr{L}^\Sigma$ by $\mathfrak{SEL}(\mathscr{L}^\Sigma)$, and we call any such model an *SEL-model*.

Definition 2.2.4. (Semantics for Standard Extensional Models) An assignment for $\mathscr{L}^\Sigma$ over an extensional model $\mathfrak{M}$ is a typed function $\mathfrak{a}:\mathsf{Var}\to D_e(\mathfrak{M})$. We denote the class of all assignments on $\mathfrak{M}$ by $\mathfrak{Assn}(\mathfrak{M})$. Given an $\mathfrak{M}\in\mathfrak{ESL}(\mathscr{L}^\Sigma)$ and an

assignment $\mathfrak{a}$ on $\mathfrak{M}$, we define the *semantic-value-in-$\mathfrak{M}$-on-$\mathfrak{a}$* function ($[\![]\!]^{\mathfrak{a}}_{\mathfrak{M}}$) for $\mathbf{t} \in \mathsf{Term}(\mathscr{L}^{\Sigma})$ as follows: $[\![\mathbf{l}]\!]^{\mathfrak{a}}_{\mathfrak{M}} = v_L(\mathbf{l})$; $[\![\mathbf{c}]\!]^{\mathfrak{a}}_{\mathfrak{M}} = v_C(\mathbf{c})$; $[\![\mathbf{x}]\!]^{\mathfrak{a}}_{\mathfrak{M}} = \mathfrak{a}(\mathbf{x})$; $[\![\mathbf{t}_o^{(\tau_1,\ldots,\tau_n\to\tau_o)}\mathbf{t}_1^{\tau_1}\cdots\mathbf{t}_n^{\tau_n}]\!]^{\mathfrak{a}}_{\mathfrak{M}} = [\![\mathbf{t}_o^{(\tau_1,\ldots,\tau_n\to\tau_o)}]\!]^{\mathfrak{a}}_{\mathfrak{M}}(\langle[\![\mathbf{t}_1^{\tau_1}]\!]^{\mathfrak{a}}_{\mathfrak{M}},\ldots,[\![\mathbf{t}_n^{\tau_n}]\!]^{\mathfrak{a}}_{\mathfrak{M}}\rangle)$; $[\![(\lambda\mathbf{x}_1^{\tau_1}\cdots\mathbf{x}_n^{\tau_n}\,.\,\mathbf{t}_o^{\tau_o})]\!]^{\mathfrak{a}}_{\mathfrak{M}}$ = the function $f \in D_{(\tau_1,\ldots,\tau_n\to\tau_o)}(\mathfrak{M})$ such that, for $a_1 \in D_{\tau_1}(\mathfrak{M}),\ldots,a_n \in D_{\tau_1}(\mathfrak{M})$, $f(\langle a_1,\ldots,a_n\rangle) = [\![\mathbf{t}_o^{\tau_o}]\!]^{\mathfrak{b}}_{\mathfrak{M}}$ where $\mathfrak{b}(\mathbf{x}_1^{\tau_1}) = a_1,\ldots,\mathfrak{b}(\mathbf{x}_n^{\tau_1}) = a_n$, and $\mathfrak{b}(\mathbf{x}) = \mathfrak{a}(\mathbf{x})$ for every $\mathbf{x} \notin \{\mathbf{x}_1^{\tau_1},\ldots,\mathbf{x}_n^{\tau_n}\}$. For $\phi \in \mathsf{Fmla}(\mathscr{L}^{\Sigma})$, $\Gamma \subseteq \mathsf{Fmla}(\mathscr{L}^{\Sigma})$, and $\mathfrak{M} \in \mathfrak{ESL}(\mathscr{L}^{\Sigma})$, we write $\mathfrak{M} \vDash_E \phi$ if $[\![\phi]\!]^{\mathfrak{a}}_{\mathfrak{M}} = 1$ for all $\mathfrak{a} \in \mathfrak{Assn}$; $\vDash_E \phi$ if, for every $\mathfrak{M} \in \mathfrak{ESL}(\mathscr{L}^{\Sigma})$, $\mathfrak{M} \vDash_E \phi$; and $\Gamma \vDash_E \phi$ if, for every $\mathfrak{M} \in \mathfrak{ESL}(\mathscr{L}^{\Sigma})$, either $\mathfrak{M} \vDash_E \phi$ or $\mathfrak{M} \nvDash_E \psi$ for some $\psi \in \Gamma$.

2.3 Non-Extensional Models

It is obvious that, given any SE-frames $\mathscr{F}_1$ and $\mathscr{F}_2$, if $D_e(\mathscr{F}_1) = D_e(\mathscr{F}_2)$, then $\mathscr{F}_1 = \mathscr{F}_2$: an SE-frame is completely determined by its type-e (first-order) domain. Furthermore, every SE-frame contains only two objects in its type-t domain. Thus SE-models are not sufficiently fine-grained to capture distinctions required for metaphysical theorizing. For instance, no SE-model can adequately represent a nontrivial metaphysical necessity operator: the operator is of syntactic type $(t \to t)$, but in every SE-model $D_{(t\to t)}$ contains only four objects (the four singulary truth functions).

It is easy enough to remedy this for modal operators, on the assumption of a fixed domain. A variant on Daniel Gallin's system $\mathsf{ML_P}$ (1975: 67–78), as implemented in Williamson (2013: 222), is the most straightforward approach (although I replace Gallin's use of $\mathscr{T}^{Or}$ with $\mathscr{T}$ and make other technical modifications):

Definition 2.3.1. (Gallin Frames) A *Gallin frame* is a triple $\mathscr{F} = \langle\mathfrak{W}, w_o, G_{\mathscr{T}}\rangle$ where $\mathfrak{W}$ is a set, $w_o \in \mathfrak{W}$, and $G_{\mathscr{T}}$ is an typed structure defined as follows: G_e is a nonempty set; $G_t = {}^{\mathfrak{W}}\{0,1\}$; and $G_{(\tau_1,\ldots,\tau_n\to\tau_o)} = {}^{\mathfrak{W}}({}^{G_{\tau_1}\times\cdots\times G_{\tau_n}}G_{\tau_o})$. We term the elements of $\mathfrak{W}$ worlds. We define the *induced SE-frame* $\hat{\mathscr{F}}$ over a Gallin frame $\mathscr{F}$ as follows: $D_e(\hat{\mathscr{F}}) = G_e(\mathscr{F})$; $D_t(\hat{\mathscr{F}}) = \{0,1\}$; $D_{(\tau_1,\ldots,\tau_n\to\tau_o)}(\hat{\mathscr{F}}) = {}^{D_{\tau_1}\times\cdots\times D_{\tau_n}}D_{\tau_o}$. We define the typed partial function $\vartheta : G_{\mathscr{T}}(\mathscr{F}) \rightsquigarrow D_{\mathscr{T}}(\mathscr{F})$ as follows: for $x \in$

$G_e, \vartheta(x) = x$; for $f \in G_t$, if f is constant (i.e., for all $w_1, w_2 \in \mathfrak{W}$, $f(w_1) = f(w_2)$), then $\vartheta(f) = f(w)$ for any w; otherwise, $\vartheta(f)$ is undefined; for $f \in G_{(\tau_1,\ldots,\tau_n\to\tau_o)}$, if f is constant (i.e., for all $w_1, w_2 \in \mathfrak{W}$, $f(w_1) = f(w_2)$) and, for every $\langle x_1,\ldots,x_n,y\rangle \in f(w)$ for any w, $\langle \vartheta(x_1),\ldots,\vartheta(x_n),\vartheta(y)\rangle$ is defined, then $\vartheta(f) = \{\langle \vartheta(x_1),\ldots,\vartheta(x_n),\vartheta(y)\rangle : \langle x_1,\ldots,x_n,y\rangle \in f(w)\}$; otherwise, $\vartheta(f)$ is undefined.

Definition 2.3.2. (Gallin Models) Gallin models are defined only for languages $\mathscr{L}^\Sigma$ where $\mathscr{L}^\Sigma$ is logically protointerpreted and Σ_C contains a type-$(t \to t)$ operator $\Box$. A Gallin protomodel for such an $\mathscr{L}^\Sigma$ is a triple $\mathfrak{M} = \langle \mathscr{F}, v_L, v_C\rangle$, where $\mathscr{F}$ is a Gallin frame, $v_L : \Sigma_{L\mathscr{T}} \to G_{\mathscr{T}}(\mathscr{F})$, and $v_C : (\Sigma_C - \{\Box\})_{\mathscr{T}} \to G_{\mathscr{T}}(\mathscr{F})$. A Gallin protomodel is logic-correct if for every $\mathbf{l} \in \Sigma_L$, $\vartheta(v_L) = \ell(\hat{\mathscr{F}})(\mathbf{l})$. A *Gallin model* for $\mathscr{L}^\Sigma$ is a logic-correct Gallin protomodel for $\mathscr{L}^\Sigma$. We denote the class of all Gallin models for $\mathscr{L}^\Sigma$ by $\mathfrak{Gal}(\mathscr{L}^\Sigma)$.

Definition 2.3.3. (Semantics for Gallin Models) An assignment for $\mathscr{L}^\Sigma$ over a Gallin model $\mathfrak{M}$ is a typed function $\mathfrak{a} : \mathsf{Var}_{\mathscr{T}} \to G_{\mathscr{T}}(\mathfrak{M})$. We denote the class of all assignments on $\mathfrak{M}$ by $\mathfrak{Assn}(\mathfrak{M})$. Given an $\mathfrak{M} \in \mathfrak{Gal}(\mathscr{L}^\Sigma)$, we define the *semantic-value-in-$\mathfrak{M}$-on-$\mathfrak{a}$-at-w function* ($[\![]\!]^{\mathfrak{a}}_{\mathfrak{M}}$) for $\mathbf{t} \in \mathsf{Term}(\mathscr{L}^\Sigma)$ as follows: $[\![\mathbf{l}^e]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = v_L(\mathbf{l})$; $[\![\mathbf{l}^\tau]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = v_L(\mathbf{l})(w)$ for $\tau \neq e$; $[\![\mathbf{c}^e]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = v_L(\mathbf{c})$; $[\![\mathbf{c}^\tau]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = v_L(\mathbf{c})(w)$ for $\tau \neq e$; $[\![\mathbf{x}^e]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = \mathfrak{a}(\mathbf{x})$; $[\![\mathbf{x}^\tau]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = \mathfrak{a}(\mathbf{x})(w)$ for $\tau \neq e$; $[\![\mathbf{t}_o^{(\tau_1,\ldots,\tau_n\to\tau_o)}\mathbf{t}_1^{\tau_1}\cdots\mathbf{t}_n^{\tau_n}]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = [\![\mathbf{t}_o^{(\tau_1,\ldots,\tau_n\to\tau_o)}]\!]^{w,\mathfrak{a}}_{\mathfrak{M}}(\langle [\![\mathbf{t}_1^{\tau_1}]\!]^{w,\mathfrak{a}}_{\mathfrak{M}},\ldots,[\![\mathbf{t}_n^{\tau_n}]\!]^{w,\mathfrak{a}}_{\mathfrak{M}}\rangle)$; $[\![(\lambda\mathbf{x}_1^{\tau_1}\cdots\mathbf{x}_n^{\tau_n}\,.\,\mathbf{t}_o^{\tau_o})]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = f(w)$ where f is the function $f \in G_{(\tau_1,\ldots,\tau_n\to\tau_o)}(\mathfrak{M})$ such that, for all $a_1 \in D_{\tau_1}(\mathfrak{M}),\ldots,a_n \in D_{\tau_1}(\mathfrak{M})$, for all $w' \in \mathfrak{W}$, $f(\langle a_1,\ldots,a_n\rangle)(w') = [\![\mathbf{t}_o^{\tau_o}]\!]^{w',\mathfrak{b}}_{\mathfrak{M}}$ where $\mathfrak{b}(\mathbf{x}_1^{\tau_1}) = a_1,\ldots,\mathfrak{b}(\mathbf{x}_n^{\tau_1}) = a_n$, and $\mathfrak{b}(\mathbf{x}) = \mathfrak{a}(\mathbf{x})$ for every $\mathbf{x} \notin \{\mathbf{x}_1^{\tau_1},\ldots,\mathbf{x}_n^{\tau_n}\}$; $[\![\Box\phi]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = 1$ if, for all $w' \in \mathfrak{W}$, $[\![\Box\phi]\!]^{w',\mathfrak{a}}_{\mathfrak{M}} = 1$ and $[\![\Box\phi]\!]^{w,\mathfrak{a}}_{\mathfrak{M}} = 0$ otherwise. We define the *actual-semantic-value-in-$\mathfrak{M}$-on-$\mathfrak{a}$ function* ($[\![]\!]^{\mathfrak{a}}_{\mathfrak{M}}$) for $\mathbf{t} \in \mathsf{Term}(\mathscr{L}^\Sigma)$ as follows: $[\![\mathbf{t}]\!]^{\mathfrak{a}}_{\mathfrak{M}} = [\![\mathbf{t}]\!]^{w_o,\mathfrak{a}}_{\mathfrak{M}}$. For $\phi \in \mathsf{Fmla}(\mathscr{L}^\Sigma)$, $\Gamma \subseteq \mathsf{Fmla}(\mathscr{L}^\Sigma)$, and $\mathfrak{M} \in \mathfrak{Gal}(\mathscr{L}^\Sigma)$, we write $\mathfrak{M} \vDash_G \phi$ if $[\![\phi]\!]^{\mathfrak{a}}_{\mathfrak{M}} = 1$ for all $\mathfrak{a} \in \mathfrak{Assn}$; $\vDash_G \phi$ if, for every $\mathfrak{M} \in \mathfrak{Gal}(\mathscr{L}^\Sigma)$, $\mathfrak{M} \vDash_G \phi$; and $\Gamma \vDash_G \phi$ if, for every $\mathfrak{M} \in \mathfrak{Gal}(\mathscr{L}^\Sigma)$, either $\mathfrak{M} \vDash_G \phi$ or $\mathfrak{M} \nvDash_G \psi$ for some $\psi \in \Gamma$.

Gallin models, as formulated here, incorporate certain simplifications compared to

standard (pointed) Kripke models for quantified first-order logic.[22] First of all, no accessibility relation has been implemented; in effect, we assume universal accessibility, so the S5 axioms are guaranteed to hold at each world of each Gallin model (and thus at the designated world of each model). Second, we make no attempt to implement variable domains: the objects, Concepts, Superconcepts and so on existing at each world are given by a single frame (albeit one somewhat more complicated than an ordinary SE-frame); thus, both the Barcan formula and the converse Barcan formula also hold. In other words, Gallin models naturally represent a *necessitist* conception of modal space.[23] The first restriction is motivated by the fact that we intend $\Box$ to be informally interpreted as metaphysical modality; plausibly, if metaphysical modality is the maximally general notion of possibility, then any non-S5 modal operator can only be an artificial limitation of the modality of interest. The second restriction is pragmatic in motivation. Representing contingentist models for higher-order modal logic, as Peter Fritz and Jeremy Goodman have shown, is exceptionally complicated (Fritz and Goodman 2016, Fritz MS 1, Fritz MS 2). If desired, we can incorporate a primitive type-$(e \to t)$ constant $E!$ into Σ_C and use it to define inner quantifiers (either for type e alone, in a system designed to combine first-order contingentism and higher-order necessitism, or for all types, in a higher-order contingentist system), thereby avoiding some of the complications.[24]

In practice, these restrictions will make relatively little difference: our main interest in Gallin models is as a stepping stone to deeper levels of hyperintensionality. Gallin models are not philosophically attractive from the point of view of a general metaphysical inquiry such as this one. The existence of worlds in $\mathfrak{W}$ in a Gallin frame outside any particular D_τ is unfaithful to the *intended* model of the type hierarchy: if worlds are objects, then they should be present in $\overline{\mathscr{D}}_e$; if, on the other hand, talk of 'worlds' is to be understood as a *façon de parler* for 'world-propositions' (Fine 1977)—maximal

[22] I adopt the terminology of 'pointed model' (a model with a designated world) from Williamson (2013: 86).

[23] See Williamson (2013) in general for the necessitism/contingentism debate.

[24] Williamson (2013: 266–76) argues against the combination of first-order contingentism and higher-order necessitism; for rebuttals, see Goodman (2016) and Fairchild (MS).

metaphysically possible propositions—then these should lie in $\overline{\mathscr{D}}_t$, along with other propositions. But Gallin frames, like standard extensional frames, admit only two entities in D_t: they represent finer-than-extensional distinctions in the semantic values of sentences, but they do so in a manner that is not perspicuously reflected in their type structure. Nonetheless, Gallin models have instrumental use: in a precise sense, every Gallin model can be converted into a model of a more general sort that we shall now develop.

The basic idea behind these more general models is to retain the use of a single typed structure as a frame, but to reject the equation of the t-type domain with $\{0,1\}$ and to decouple the domains for derived types from sets of functions built up over more basic types. The most general method of doing this is due to Reinhard Muskens (2007). On a Muskens-style approach, instead of a single typed collection $D_{\mathscr{T}}$, we have two: a primary collection $I_{\mathscr{T}}$ of *intensions* and a derived collection $E_{\mathscr{T}}$ of *extensions*. For basic types, $I_e = E_e$, corresponding to the principle that there is no genuinely worldly hyperintensional behaviour at the level of *objects*. The type E_t is $\{0,1\}$, but I_t has no size restrictions at all: the general characterization of a frame of this type leaves the details of the space of propositions wholly open. Similarly, for every complex type $(\tau_1,\ldots,\tau_n \to \tau_0)$, the only constraint on $I_{(\tau_1,\ldots,\tau_n\to\tau_0)}$ is that it be disjoint from $I_{\tau'}$ for any distinct type τ': $E_{(\tau_1,\ldots,\tau_n\to\tau_0)}$, however, is the set of (set-theoretic) functions from $I_{\tau_1},\ldots,I_{\tau_n}$ to I_{τ_0}. In other words, where τ is a complex type, the elements of I_τ represent the *functions-in-intension* of type τ, which are not taken to have internal structure; the elements of E_τ represent set-theoretical *functions-in-extension* constructed over functions-in-intension of other less complex types. As Muskens writes,

> The aim of this paper is not to add one more theory of intension to the proposals that have already been made, but is an investigation of their common underlying logic. The idea will be that the two-stage set-up is essentially *all* that is needed to obtain intensionality. For the purposes of logic it suffices to consider intensions as abstract objects; the question what intensions *are*,

while philosophically important, can be abstracted from. (Muskens 2007: 101)

A typed extension function η links elements of I_τ to those of E_τ: η is the identity for $\tau = e$ but is otherwise unconstrained. Crucially, in general there can be distinct $a_1, a_2 \in I_\tau$ such that $\eta(a_1) = \eta(a_2)$. It is this that allows for the extreme generality of hyperintensional fineness of grain permitted by Muskens's approach: there are frames with no more than one extension for every intension, but there are also frames containing an extension corresponding to fifty or $\aleph_2$ or the least measurable cardinal-many intensions.

More formally, we define a Muskens-style system, modified to use $\mathscr{T}$ rather than $\mathscr{T}^{Or}$ and to cohere with our general notion of logic-correctness, as follows:[25]

Definition 2.3.4. (Muskens Frames) A *Muskens protoframe (M-protoframe)* is a typed collection $I_\mathscr{T}$ where I_e is nonempty, I_t contains at least two elements, and $I_{\tau_1} \cap I_{\tau_2} = \emptyset$ when $\tau_1 \neq \tau_2$. A *Muskens frame (M-frame)* is a pair $\mathscr{F} = \langle I_\mathscr{T}, \eta \rangle$, where $I_\mathscr{T}$ is an M-protoframe and η is defined as follows. Let $E_e = I_e; E_t = \{0, 1\}$; $E_{(\tau_1,\ldots,\tau_n \to \tau_0)} = {}^{I_{\tau_1} \times \cdots \times I_{\tau_n}} I_{\tau_0}$; $\eta : I \to E$ is a typed function under the constraint that η_e is the identity and there exist $x_1, x_2 \in I_t$ such that $\eta(x_1) = 0$ and $\eta(x_1) = 1$. We define $I = \bigcup_{\tau \in \mathscr{T}} I_\tau$. An M-frame is *full* if, for every $\tau \in \mathscr{T}$, for every $x \in E_\tau$, there exists some $y \in I_\tau$ such that $x = \eta(y)$.

In addition to the *extension function* η, it will be useful to define the *hereditary extension partial function* ξ, which delivers the set-theoretical construction from E_t and E_e that results from repeated substitution of extensions for intensions. A hereditary extension is well-defined in the case of complex types when no function-in-intension at any stage differentiates among two extensionally equivalent arguments—in other words, for entities that are 'extensional all the way down'.

We define the *extensional kernel* $K_\mathscr{T}(\mathscr{F})$ of an M-frame $\mathscr{F}$ as follows: $K_\mathscr{T}(\mathscr{F}) = \{K_\tau(\mathscr{F}) : \tau \in \mathscr{T}\}$ where $K_e(\mathscr{F}) = E_e(\mathscr{F})$; $K_{\langle\rangle}(\mathscr{F}) = E_{\langle\rangle}(\mathscr{F})$; $K_{\langle\tau_1,\ldots,\tau_n\rangle}(\mathscr{F}) =$

[25] Other inessential changes of presentation, notation, and terminology are introduced, here and in the presentation of BBK-models below, to make comparisons between different frameworks easier.

$^{K_{\tau_1}(\mathscr{F})\times\cdots\times K_{\tau_n}(\mathscr{F})}K_{\langle\rangle}(\mathscr{F})$. We now define the typed partial function $\xi : I(\mathscr{F}) \rightsquigarrow K(\mathscr{F})$ recursively as follows: if $x \in I_e \cup I_t$, $\xi(x) = x$; if $f \in I_{(\tau_1,\ldots,\tau_n\to\tau_o)}$, then $\xi(f) = \{\langle\xi(x_1),\ldots,\xi(x_n),\xi(y)\rangle : \langle x_1,\ldots x_n,y\rangle \in f\}$ if each of $\{\xi(x_1),\ldots,\xi(x_n),\xi(y)\}$ is defined and $\{\langle\xi(x_1),\ldots,\xi(x_n),\xi(y)\rangle : \langle x_1,\ldots x_n,y\rangle \in f\}$ is functional; it is undefined otherwise.

Definition 2.3.5. (Muskens Models) Let $\mathscr{L}^\Sigma$ be a logically protointerpreted language. A *variable assignment for $\mathscr{L}^\Sigma$ on the M-frame $\mathscr{F}$* is a typed function $\mathfrak{a} : \mathsf{Var} \to I$. We denote the class of all variable assignments for $\mathscr{L}^\Sigma$ on $\mathscr{F}$ by $\mathfrak{Assn}$. An *interpretation function* v for $\mathscr{L}^\Sigma$ on $\mathscr{F}$ is a typed function $v : \mathfrak{Assn} \times \mathsf{Term}(\mathscr{L}^\Sigma) \to I$ (i.e., in this context, a function such that, if $\mathbf{t} \in \mathsf{Term}_\tau$, then $v(\mathfrak{a},\mathbf{t}) \in I_\tau$ for all $\mathfrak{a} \in \mathfrak{Assn}$) subject to the following constraints: (1) if $\mathbf{x} \in \mathsf{Var}$, $v(\mathfrak{a},\mathbf{x}) = \mathfrak{a}(x)$; (2) for all $\mathfrak{a},\mathfrak{b} \in \mathfrak{Assn}$, if for every $\mathbf{x}$ free in $\mathbf{t}$, $\mathfrak{a}(\mathbf{x}) = \mathfrak{b}(\mathbf{x})$, then $v(\mathfrak{a},\mathbf{t}) = v(\mathfrak{b},\mathbf{t})$; (3) $v(\mathfrak{a},\mathbf{t}[\mathbf{t}_o/\mathbf{x}]) = v(\mathfrak{a}',\mathbf{t})$ where $\mathfrak{a}'(\mathbf{x}) = v(\mathfrak{a},\mathbf{t}_o)$ and $\mathfrak{a}'(\mathbf{y}) = \mathfrak{a}(\mathbf{y})$ for all $\mathbf{y} \neq \mathbf{x}$. A *Muskens protomodel* for $\mathscr{L}^\Sigma$ is a pair $\mathfrak{M} = \langle\mathscr{F},v\rangle$ where $\mathscr{F}$ is an M-frame and v is an interpretation function for $\mathscr{L}^\Sigma$ on $\mathscr{F}$. $\mathfrak{M}$ is full just in case $\mathscr{F}$ is. Holding fixed a particular $\mathfrak{M}$, we define the typed function $\tilde{v} : \mathfrak{Assn} \times \mathsf{Term}(\mathscr{L}^\Sigma) \to I$ as $\langle\mathfrak{a},x\rangle \mapsto \eta(v(\mathfrak{a},x))$. $\mathfrak{M}$ is *logic-correct* just in case (1) for every $\mathbf{l} \in \Sigma_L$ and every $\mathfrak{a} \in \mathsf{Assn}$, $\xi(v(\mathfrak{a},\mathbf{l})) = \ell(\mathbf{l})$; (2) if $\mathbf{t}_o \in \mathsf{Term}_{(\tau_1,\ldots,\tau_n\to\tau_o)}, \mathbf{t}_1 \in \mathsf{Term}_{\tau_1},\ldots,\mathbf{t}_n \in \mathsf{Term}_{\tau_1}$, then for all $\mathfrak{a}$, $v(\mathfrak{a},\mathbf{t}_o\mathbf{t}_1\cdots\mathbf{t}_n) = \tilde{v}(\mathfrak{a},\mathbf{t}_o)(\langle v(\mathfrak{a},\mathbf{t}_1),\ldots,v(\mathfrak{a},\mathbf{t}_1)\rangle)$; (3) if $\mathbf{x}_1 \in \mathsf{Var}_{\tau_1},\ldots,\mathbf{x}_n \in \mathsf{Var}_{\tau_n}$, and $\mathbf{t} \in \mathsf{Term}_{\tau_o}$, then $\tilde{v}(\mathfrak{a},(\lambda\mathbf{x}_1\cdots\mathbf{x}_n\,.\,\mathbf{t}))$ is the function f such that, for all $a_1 \in I_{\tau_1},\ldots,a_n \in I_{\tau_n}$, $f(a_1,\ldots,a_n) = v(\mathfrak{a}',\mathbf{t})$ where $\mathfrak{a}'(\mathbf{x}_1) = a_1,\ldots,\mathfrak{a}'(\mathbf{x}_n) = a_n$, and $\mathfrak{a}'(\mathbf{y}) = \mathfrak{a}(\mathbf{y})$ for all $\mathbf{y} \notin \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. For $\star \subseteq \{\alpha,\beta,\gamma\}$, using the notation of Definition 2.2.1, $\mathfrak{M}$ is *intensionally $\star$-respecting* just in case, for all $\mathfrak{a} \in \mathfrak{Assn}$ and all $\mathbf{t}_1,\mathbf{t}_2 \in \mathsf{Term}$, if $\mathbf{t}_1 \approx_\star \mathbf{t}_2$, $v(\mathfrak{a},\mathbf{t}_1) = v(\mathfrak{a},\mathbf{t}_2)$; $\mathfrak{M}$ is *extensionally $\star$-respecting* just in case, for all $\mathfrak{a} \in \mathfrak{Assn}$ and all $\mathbf{t}_1,\mathbf{t}_2 \in \mathsf{Term}$, if $\mathbf{t}_1 \approx_\star \mathbf{t}_2$, $\tilde{v}(\mathfrak{a},\mathbf{t}_1) = \tilde{v}(\mathfrak{a},\mathbf{t}_2)$. A *Muskens model* (M-model) is a full, logic-correct, intensionally $\alpha\beta\eta$-respecting Muskens protomodel. We denote the class of all M-models for $\mathscr{L}^\Sigma$ by $\mathfrak{Msk}(\mathscr{L}^\Sigma)$.

Definition 2.3.6. (Semantics for Muskens Models) For a $\phi \in \mathsf{Fmla}(\mathscr{L}^\Sigma)$ and $\mathfrak{M} \in$

$\mathfrak{Mdl}(\mathscr{L}^{\Sigma})$, if $\eta(v(\mathfrak{a},\phi)) = 1$ for all $\mathfrak{a} \in \mathfrak{Assn}$, we write $\mathfrak{M} \vDash_M \phi$. If $\mathfrak{M} \vDash_M \phi$ for all $\mathfrak{M} \in \mathfrak{Mdl}(\mathscr{L}^{\Sigma})$, we write $\vDash_M \phi$; if, for every $\mathfrak{M} \in \mathfrak{Mdl}(\mathscr{L}^{\Sigma})$, either $\mathfrak{M} \vDash \phi$ or $\mathfrak{M} \nvDash \psi$ for some $\psi \in \Gamma$, we write $\Gamma \vDash_M \phi$.

The decision to restrict consideration to intensionally $\alpha\beta\eta$-respecting models, reflecting a robust commitment to classical logic, will be defended below.

The framework developed by Benzmüller, Brown, and Kohlhase (2004), like that of Muskens, starts by setting aside the question of what intensions are. In fact, we can simply use M-protoframes as the basis for BBK-models. (In their own presentation, Benzmüller, Brown, and Kohlhase (2004: 1035) do not require their typed domains to be pairwise disjoint, but adding the requirement makes no practical difference.) Instead of adding an extension function to an M-protoframe to form a frame, however, Benzmüller, Brown, and Kohlhase construct *applicative structures*, in which only the members of the type-t domain have 'extensions' (truth values) and all other cross-domain relationships are handled by a separate function application operator. A version of their approach, modified to use $\mathscr{T}$ rather than $\mathscr{T}^{Ch}$ and to incorporate logic-correctness, proceeds as follows:

Definition 2.3.7. (BBK-Applicative Structures) Let $\mathscr{T}^{Cpx} = \mathscr{T} - \{e,t\}$; thus, for $\tau \in \mathscr{T}^{Cpx}$, $\tau = (\tau_1,\ldots,\tau_n \to \tau_o)$ where $\tau_o,\ldots,\tau_n \in \mathscr{T}$. Where $I_{\mathscr{T}}$ is an M-protoframe, for each $\tau \in \mathscr{T}^{Cpx}$, an *application operator for* τ on $I_{\mathscr{T}}$ is a function $\nabla_{(\tau_1,\ldots,\tau_n\to\tau_o)} \in {}^{(I_{(\tau_1,\ldots,\tau_n\to\tau_o)}\times I_{\tau_1}\times\cdots\times I_{\tau_n})}I_{\tau_o}$. An *application family* on $I_{\mathscr{T}}$ is a typed collection $\mathring{\nabla} = \{\nabla_\tau : \tau \in \mathscr{T}^{Cpx}\}$. We write the result of applying ∇_τ to $(x,y_1,\ldots,y_n)$, when these are appropriately typed, using infix notation as $x\nabla_\tau y_1\cdots y_n$, and we leave off the subscript τ when no confusion will result. A BBK-applicative structure (BBK-AS) is a pair $\mathscr{F} = \langle I_{\mathscr{T}}, \mathring{\nabla}\rangle$ where $I_{\mathscr{T}}$ is an M-protoframe and $\mathring{\nabla}$ is an application family on $I_{\mathscr{T}}$.

Definition 2.3.8. (BBK-Protomodels) Let $\mathscr{L}^{\Sigma}$ be a logically protointerpreted language. A *variable assignment for* $\mathscr{L}^{\Sigma}$ on the BBK-AS $\mathscr{F}$ is a typed function $\mathfrak{a}$: Var $\to I$. We denote the class of all variable assignments for $\mathscr{L}^{\Sigma}$ on $\mathscr{F}$ by

$\mathfrak{Assn}$. An *interpretation function* v for $\mathscr{L}^\Sigma$ on $\mathscr{F}$ is a typed function $v : \mathfrak{Assn} \times \mathsf{Term}(\mathscr{L}^\Sigma) \to I$ (i.e., in this context, a function such that, if $\mathbf{t} \in \mathsf{Term}_\tau$, then $v(\mathfrak{a},\mathbf{t}) \in I_\tau$ for all $\mathfrak{a} \in \mathfrak{Assn}$) subject to the following constraints: (1) if $\mathbf{x} \in \mathsf{Var}$, $v(\mathfrak{a},\mathbf{x}) = \mathfrak{a}(x)$; (2) for all $\mathfrak{a},\mathfrak{b} \in \mathfrak{Assn}$, if for every $\mathbf{x}$ free in $\mathbf{t}$, $\mathfrak{a}(\mathbf{x}) = \mathfrak{b}(\mathbf{x})$, then $v(\mathfrak{a},\mathbf{t}) = v(\mathfrak{b},\mathbf{t})$; (3) $v(\mathfrak{a},\mathbf{t}[\mathbf{t}_0/\mathbf{x}]) = v(\mathfrak{a}',\mathbf{t})$ where $\mathfrak{a}'(\mathbf{x}) = v(\mathfrak{a},\mathbf{t}_0)$ and $\mathfrak{a}'(\mathbf{y}) = \mathfrak{a}(\mathbf{y})$ for all $\mathbf{y} \neq \mathbf{x}$. A *truth-mapping* on the BBK-AS $\mathscr{F}$ is a function $\Theta : D_t \to \{0,1\}$. A BBK-protomodel is a triple $\mathfrak{M} = \langle \mathscr{F}, v, \Theta \rangle$, where $\mathscr{F}$ is a BBK-AS, v is an interpretation function for $\mathscr{L}^\Sigma$ on $\mathscr{F}$, and Θ is a truth-mapping on $\mathscr{F}$.

We set $\hat{E}_e = I_e; \hat{E}_t = \{0,1\}; \hat{E}_{(\tau_1,\ldots,\tau_n\to\tau_0)} = {}^{\hat{E}_{\tau_1}\times\cdots\times\hat{E}_{\tau_n}}\hat{E}_{\tau_0}$, and we define the *hereditary quasi-extension* typed partial function $\xi : I \rightsquigarrow \hat{E}$ as follows: for $x \in I_t$, $\xi(x) = \Theta(x)$; for $x \in I_e$, $\xi(x) = x$; for $x \in I_{(\tau_1,\ldots,\tau_n\to\tau_0)}$, we define $g(x) \in {}^{I_{\tau_1}\times\cdots\times I_{\tau_n}}I_{\tau_0}$ as the function $\langle y_1,\ldots,y_n\rangle \mapsto x\nabla_{(\tau_1,\ldots,\tau_n\to\tau_0)}y_1\cdots y_n$ and set $\xi(x) = \{\langle \xi(a_1),\ldots,\xi(a_n),\xi(a_0)\rangle : \langle a_1,\ldots, a_n,a_0\rangle \in g(x)\}$ if all of $\{\xi(a_1),\ldots,\xi(a_n),\xi(a_0)\}$ are defined and $\{\langle \xi(a_1),\ldots, \xi(a_n),\xi(a_0)\rangle : \langle a_1,\ldots,a_n,a_0\rangle \in g(x)\}$ is functional; $\xi(x)$ is undefined otherwise.

Definition 2.3.9. (BBK-Models) A BBK-protomodel is *full* just in case, for every $x \in \hat{E}_\tau$, there exists some $y \in I_\tau$ such that $\xi(x) = y$. A BBK-protomodel is *logic-correct* just in case (1) for every $\mathbf{l} \in \Sigma_L$ and every $\mathfrak{a} \in \mathsf{Assn}$, $\xi(v(\mathfrak{a},\mathbf{l})) = \ell(\mathbf{l})$; (2) if $\mathbf{t}_0 \in \mathsf{Term}_{(\tau_1,\ldots,\tau_n\to\tau_0)}, \mathbf{t}_1 \in \mathsf{Term}_{\tau_1},\ldots,\mathbf{t}_n \in \mathsf{Term}_{\tau_1}$, then for all $\mathfrak{a}$, $v(\mathfrak{a},\mathbf{t}_0\mathbf{t}_1\cdots\mathbf{t}_n) = v(\mathfrak{a},\mathbf{t}_0)\nabla_{(\tau_1,\ldots,\tau_n\to\tau_0)}(\langle v(\mathfrak{a},\mathbf{t}_1),\ldots,v(\mathfrak{a},\mathbf{t}_1)\rangle)$; (3) if $\mathbf{x}_1 \in \mathsf{Var}_{\tau_1},\ldots,\mathbf{x}_n \in \mathsf{Var}_{\tau_n}$, and $\mathbf{t} \in \mathsf{Term}_{\tau_0}$, then $v(\mathfrak{a},(\lambda\mathbf{x}_1\cdots\mathbf{x}_n\,.\,\mathbf{t}))$ is the function f such that, for all $a_1 \in I_{\tau_1},\ldots,a_n \in I_{\tau_n}$, $f(a_1,\ldots,a_n) = v(\mathfrak{a}',\mathbf{t})$ where $\mathfrak{a}'(\mathbf{x}_1) = a_1,\ldots,\mathfrak{a}'(\mathbf{x}_n) = a_n$, and $\mathfrak{a}'(\mathbf{y}) = \mathfrak{a}(\mathbf{y})$ for all $\mathbf{y} \notin \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. For $\star \subseteq \{\alpha,\beta,\gamma\}$, using the notation of Definition 2.2.1, $\mathfrak{M}$ is *intensionally $\star$-respecting* just in case, for all $\mathfrak{a} \in \mathfrak{Assn}$ and all $\mathbf{t}_1,\mathbf{t}_2 \in \mathsf{Term}$, if $\mathbf{t}_1 \approx_\star \mathbf{t}_2$, $v(\mathfrak{a},\mathbf{t}_1) = v(\mathfrak{a},\mathbf{t}_2)$; $\mathfrak{M}$ is *extensionally $\star$-respecting* just in case, for all $\mathfrak{a} \in \mathfrak{Assn}$ and all $\mathbf{t}_1,\mathbf{t}_2 \in \mathsf{Term}$, if $\mathbf{t}_1 \approx_\star \mathbf{t}_2$, $\xi(v(\mathfrak{a},\mathbf{t}_1)) = \xi(v(\mathfrak{a},\mathbf{t}_2))$. A *BBK-model* is a full, logic-correct, intensionally $\alpha\beta\eta$-respecting BBK-protomodel. We denote the class of all BBK-models for $\mathscr{L}^\Sigma$ by $\mathfrak{BBK}(\mathscr{L}^\Sigma)$.

Definition 2.3.10. (Semantics for BBK-Models) For a $\phi \in \mathsf{Fmla}(\mathscr{L}^\Sigma)$ and $\mathfrak{M} \in$

$\mathfrak{BBK}(\mathscr{L}^{\Sigma})$, if $\Theta(v(\mathfrak{a},\phi))$ for all $\mathfrak{a} \in \mathfrak{Assn}$, we write $\mathfrak{M} \vDash_{BBK} \phi$. If $\mathfrak{M} \vDash_{BBK} \phi$ for all $\mathfrak{M} \in \mathfrak{BBK}(\mathscr{L}^{\Sigma})$, we write $\vDash_{BBK} \phi$; if, for every $\mathfrak{M} \in \mathfrak{BBK}(\mathscr{L}^{\Sigma})$, either $\mathfrak{M} \vDash \phi$ or $\mathfrak{M} \vDash \psi$ for some $\psi \in \Gamma$, we write $\Gamma \vDash_{BBK} \phi$.

In contrast to Muskens, Benzmüller, Brown, and Kohlhase offer very little in the way of intuitive justification for the choices involved in their framework, and *prima facie* it is difficult not to concur with Muskens's comment that "such non-standard evaluations [by way of the ∇ application operators] seem unnecessary" (Muskens 2007: 99). But it may be possible to motivate the system of BBK-applicative structures on the basis of a certain picture about the relation between the semantic value of terms and the fineness of grain in reality.[26] On this view, taking the distinction between functions-in-intension and functions-in-extension metaphysically seriously—as is required by the picture behind M-models—is a mistake: *terms* have extensions and intensions, but there is no structure *at the level of fundamental reality* that parallels the determination of a term's reference by its sense. Higher-order entities do not come equipped with intrinsically determined extensions of the sort that are naturally amenable to representation in set theory; rather, there is merely the direct application of one entity to another, unmediated by constraints to ensure that functions-in-extension match up with functions-in-intension: all the work of coordination is carried out by the ∇ application operators.

This picture is not without a certain appeal, but it ultimately leads to more metaphysical difficulties than it resolves. First, the rejection of extensions leaves the role of the truth-mapping Θ in the BBK-model utterly mysterious: why should type-$\bar{t}$ entities alone have a special connexion to something outside the BBK-applicative structure? If the answer is merely that a model theory is useless unless it connects up in some way to truth and falsity of sentences, then this leads to a quandry. There are many kinds of model theories for formal languages that are deployed merely instrumentally, but the non-extensional semantics for type theory used in a foundational investiga-

[26] I owe the following defence of Benzmüller, Brown, and Kohlhase to Cian Dorr, in conversation.

tion should not be one of them: we want a model theory that captures at least some important features of the intended interpretation of the framework. The requirement that type-$\overline{t}$ entities connect up somehow to truth and falsity does not lessen when we are using model theory in a foundational context; if anything, it becomes even more pressing. Thus we have some evidence that this requirement reflects a genuine feature of propositions. But a framework that admits of extensions for *some types but not others* is significantly less uniform than one which admits extensions for all types across the board.

There is a second, perhaps more serious, problem with this attempt to motivate Benzmüller, Brown, and Kohlhase's framework. Each $\nabla_{(\tau_1,\ldots,\tau_n\to\tau_o)}$ function application operator itself acts on (representatives in the model theory of) type-$\overline{(\tau_1,\ldots,\tau_n\to\tau_o)}$, type-$\overline{\tau_1}$, …, type-$\overline{\tau_n}$ entities to yield a (representative of a) type-$\overline{\tau_o}$ entity. Thus, it should be representable within the theory itself as a type-$\overline{((\tau_1,\ldots,\tau_n\to\tau_o),\tau_1,\ldots,\overline{\tau_n\to\tau_o})}$ entity—which would then require a $\nabla_{((\tau_1,\ldots,\tau_n\to\tau_o),\tau_1,\ldots,\tau_n\to\tau_o)}$ application operator, and so on: Bradley's regress threatens.

At this point, the proponent of BBK-models may offer a *tu quoque* objection: the same argument can be run for ordinary function application, even in purely extensional settings. In standard mathematics (and in the account of type theory using M-models), we escape the regress by simply treating the function application 'operator' as syncategorematic: in effect, function application is not truly an *operation* at all, for (like the λ variable binder) it corresponds to nothing at any level of the type hierarchy. Why is it not open to the defender of the BBK picture to do the same? There is a significant difference, however, inasmuch as the syncategorematic notion of function application that the defender of M-models needs is the familiar one that is required in any theory that displays a minimum of type stratification. The defender of BBK-models needs this notion, too—in the metatheory in which he constructs his set-theoretic models—but he also needs the ∇ operators, which appear nowhere else.

In any case, a philosopher who wishes to investigate model classes with an eye towards faithfulness need not worry that she is forgoing any useful resources by using

M-models instead of BBK-models: any BBK-model can be converted into an M-model that makes the same formulas true. Before showing this, we show that M-models similarly generalize Gallin models:

Theorem 2.3.11. If $\mathfrak{M} \in \mathfrak{Gal}(\mathscr{L}^{\Sigma})$, then there exists an $\mathfrak{N} \in \mathfrak{Msk}(\mathscr{L}^{\Sigma})$ such that $\mathfrak{M} \vDash \phi$ if and only if $\mathfrak{N} \vDash \phi$.[27]

Proof. (Sketch.) To reduce clutter, we use superscripts as indices for the models. We set $I^{\mathfrak{N}}_{\tau} = G^{\mathfrak{M}}_{\tau}$; we define $\eta^{\mathfrak{N}}$ as follows: for $x \in I^{\mathfrak{N}}_{e}$, $\eta(x) = x$; for $f \in I^{\mathfrak{N}}_{\tau}$, $\eta(f) = 1$ if $f(w_0) = 1$ and $\eta(f) = 0$ otherwise; for all other types, $\eta(f) = f(w_0)$. Define $v^{\mathfrak{N}} : \mathfrak{Assn} \times \mathsf{Term}(\mathscr{L}^{\Sigma}) \to I$ as follows: (1) for $\mathbf{x} \in \mathsf{Var}$, $v^{\mathfrak{N}}(\mathbf{x}, \mathfrak{a}) = \mathfrak{a}(x)$; (2) for $\mathbf{t} \in \Sigma_C \cup \Sigma_L - \{\Box\}$, $v^{\mathfrak{N}}(\mathfrak{a}, \mathbf{t}) = v^{\mathfrak{M}}(\mathbf{t})$; (3) define $\top \in I^{\mathfrak{N}}_{t}$ as the constant function f such that $f(w) = 1$ for each $w \in \mathfrak{W}$; let $\bot \in I^{\mathfrak{M}}_{t}$ be the constant function such that $f(w) = 0$ for each $w \in \mathfrak{W}$; let $v^{\mathfrak{N}}(\mathfrak{a}, \Box)$ be the constant function f such that $f(w) = g$ for each $w \in \mathfrak{W}$, where $g \in E_{(w \to w)}$ is the function such that $g(\top) = \top$ and $g(x) = \bot$ for $x \neq \top$; (4) if $\mathbf{t}_0 \in \mathsf{Term}_{(\tau_1,\ldots,\tau_n \to \tau_0)}, \mathbf{t}_1 \in \mathsf{Term}_{\tau_1}, \ldots, \mathbf{t}_n \in \mathsf{Term}_{\tau_1}$, then for all $\mathfrak{a}$, $v^{\mathfrak{N}}(\mathfrak{a}, \mathbf{t}_0\mathbf{t}_1 \cdots \mathbf{t}_n) = \eta^{\mathfrak{N}}(v^{\mathfrak{N}}((\mathfrak{a}, \mathbf{t}_0))(\langle v^{\mathfrak{N}}(\mathfrak{a}, \mathbf{t}_1), \ldots, v^{\mathfrak{N}}(\mathfrak{a}, \mathbf{t}_1)\rangle))$; and (5) if $\mathbf{x}_1 \in \mathsf{Var}_{\tau_1}, \ldots, \mathbf{x}_n \in \mathsf{Var}_{\tau_n}$, and $\mathbf{t} \in \mathsf{Term}_{\tau_0}$, then $v^{\mathfrak{N}}(\mathfrak{a}, (\lambda \mathbf{x}_1 \cdots \mathbf{x}_n \,.\, \mathbf{t}))$ is the constant function f such that $f(w) = g$ for each $w \in \mathfrak{W}$, where, for all $a_1 \in I_{\tau_1}, \ldots, a_n \in I_{\tau_n}$, $f(a_1, \ldots, a_n) = v(\mathfrak{a}', \mathbf{t})$ where $\mathfrak{a}'(\mathbf{x}_1) = a_1, \ldots, \mathfrak{a}'(\mathbf{x}_n) = a_n$, and $\mathfrak{a}'(\mathbf{y}) = \mathfrak{a}(\mathbf{y})$ for all $\mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. It is immediate that this is an interpretation function. That $\mathfrak{M} \vDash \phi$ implies $\mathfrak{N} \vDash \phi$ follows from these definitions by a straightforward induction on the complexity of formulas. ⊣

We now prove the main result.

Theorem 2.3.12. If $\mathfrak{M} \in \mathfrak{BBK}(\mathscr{L}^{\Sigma})$ and $\mathfrak{M} \vDash \phi$, then there exists an $\mathfrak{N} \in \mathfrak{Msk}(\mathscr{L}^{\Sigma})$ such that $\mathfrak{N} \vDash \phi$.

Proof. (Sketch) We set $I^{\mathfrak{N}}_{\tau} = I^{\mathfrak{M}}_{\tau}$. We define $\eta^{\mathfrak{N}}$ as follows: for $x \in I^{\mathfrak{N}}_{t}$, $\eta^{\mathfrak{N}}(x) = \Theta^{\mathfrak{M}}(x)$; for $x \in I^{\mathfrak{N}}_{e}$, η is the identity; for $x \in I^{\mathfrak{N}}_{(\tau_1,\ldots,\tau_n \to \tau_0)}$, $\eta(x) =$ the function $z_1, \ldots, z_n \mapsto$

[27] This result is suggested by, although more general than, the construction given by Muskens (2007: 114–17).

$x\nabla_{(\tau_1,\ldots,\tau_n\to\tau_o)}(z_1,\ldots,z_n)$. We simply let $v^{\mathfrak{N}} = v^{\mathfrak{M}}$. The crucial step is verifying that this is in fact an interpretation function; in particular, we note that clause (2) of the requirement is satisfied as a result of the fact that the quasi-extension function $\xi^{\mathfrak{N}}$ incorporates ∇-application. The remainder of the proof is, once again, an induction on the complexity of formulas. ⊣

Since BBK-models can be transformed into M-models, any system of hyperintensional metaphysics developed over the former can be imported into the latter. This is important, because the most systematic work on the subject—Dorr's theory of statements of the form 'to be *F* is to be *G*' (2016)—uses BBK-models.

Dorr views the locution (*e.g.*, 'to be a vixen is to be a female fox', 'to be water is to be H_2O', 'to be happy is to be not not happy', and so on) as the standard means for expressing higher-order identity theses in natural language. He focusses particularly on what, in our terms, are claims of the form $\mathbf{p} \equiv \mathbf{q}$ and $(\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}\,.\,\phi) \equiv (\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}\,.\,\psi)$ and develops several different candidate theories of *fineness of grain* for propositions and properties (Concepts, in our idiom) to handle these cases. The two that Dorr considers most promising are *Booleanism* and *Non-Circularity*. On the Boolean picture, "the truth-functional operators [for Dorr, only conjunction, disjunction, and negation] conform to the axioms of a Boolean algebra, with identification [*i.e.*, the higher-order $\equiv$] playing the role of identity" (Dorr 2016: 67): the schema $(\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}\,.\,\phi) \equiv (\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}\,.\,\psi)$ holds just in case $\phi \leftrightarrow \psi$ is a substitution instance of a propositional tautology.[28] The non-circular picture, in contrast, starts from the idea that there is something fundamentally problematic with "the term on one side of an identification occurring as a *proper constituent* of the term on the other side" (2016: 73): appropriately weakened so as not to conflict with nonvacuous β-conversion, this leads him to reject the idempotence principle $(\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}.\phi) \equiv (\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}.\phi\wedge\phi)$ (2016: 77). But other options are possible. For instance, with a privileged metaphysical modality operator, one can introduce a very coarse-grained system on which $(\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}\,.\,\phi) \equiv (\lambda\mathbf{x}^{\tau_1}\cdots\mathbf{x}^{\tau_n}\,.\,\psi)$

[28] Here $\phi \equiv \psi$ is counted as the degenerate (zero-variable) case of the schema.

just in case (the universal closure of) $\Box(\phi \leftrightarrow \psi)$ holds. Or, following Goodman (MS), one can develop a notion of *aboutness* on which a proposition essentially is *about* certain objects: in the simplest case, $\mathbf{p} \equiv \mathbf{q}$ just in case $\Box(\mathbf{p} \leftrightarrow \mathbf{q})$ and the class $\mathscr{C}(\mathbf{p})$ of objects that $\mathbf{p}$ is about is identical to the class $\mathscr{C}(\mathbf{q})$ of objects that $\mathbf{q}$ is about (with added refinements in higher types).

The example $(\lambda x \,.\, x = 1 \wedge x = 7) \neq (\lambda x \,.\, x = 2 \wedge x = 4)$, which I used to motivate consideration of hyperintensional models, rules out the coarsest-grained of these suggestions. In general, however, I can remain fairly neutral about issues of fineness of grain. My use of grounding operators in §5.2, if fully formalized, would require a fairly fine-grained approach. Rejecting the idempotence principle brings one close to the level of fineness of grain needed to introduce a grounding operator, although Dorr's particular set of theoretical choices do not comport naturally with standard principles of the logic of ground. I consider only two specific claims about fineness of grain. In §3.3, I argue that a trivialist thesis, due to Agustín Rayo, about higher-order identity involving abstraction operators is implausible. In §6.3, I discuss (but do not endorse) the NONINFLATION thesis—but this is a pure cardinality claim compatible with many detailed metaphysical pictures. The crucial point here is that Muskens models are sufficiently general that, using Theorem 2.3.12, each of Dorr's metaphysical options—ranging from the most permissive to the most specific—can be associated with an appropriate model class $\mathscr{M} \subsetneq \mathfrak{Mus}(\mathscr{L}^{\Sigma})$; commitment to the claim can be expressed by endorsing the claim that $\mathscr{M}$ is weakly faithful to truth on $\vDash_M$. Provided that we are confident that $\mathfrak{Mus}(\mathscr{L}^{\Sigma})$ is itself weakly faithful to truth on $\vDash_M$, Muskens models provide the expressively rich slate we need for metaphysical theorizing.

Should we, then, accept that $\mathfrak{Mus}(\mathscr{L}^{\Sigma})$ is weakly faithful to truth on $\vDash_M$ (indeed, weakly faithful to logical truth, on the assumption that every constant in Σ_L is genuinely logical)? The potentially controversial elements of the picture are the rejection of non-full models and the requirement that M-models be intensionally $\alpha\beta\eta$-respecting. The first was defended in §2.1, on the basis of considerations about plurals and the generality of the higher-order quantifier. As for the latter requirement, it is

crucial to note that we are assuming that failures of substitutitivity in propositional attitude contexts are not genuinely worldly and thus need not be represented within a metaphysically perspicuous system: it is for this reason that we rejected the addition of types for senses in §2.1. In any case, intensional α-equivalence is so intuitively appealing that it has been argued that it holds even in belief contexts, as in Church's account of synonymous isomorphism (1954); it is also permitted in the finest-grained formulation (Alternative 0) of LSD (Church 1946, 1951a). Intensional η-equivalence is similarly difficult to reject (Dorr 2016: 65; Salmon 2010: 454 fn. 19). As for intensional β-equivalence, as Dorr notes (2016: 52–53), it is derivable from full nonvacuous β-conversion. Leaving aside belief contexts, as he notes, the obvious counterexamples involve vacuous λ-binding (which we have excluded, as we work in a $\lambda\mathbf{I}$ language) and some extremely controversial theses on the relation of grounding to λ operators (Dorr 2016: 52–55). Not only are those theses not required for the use we make of grounding in Chapter V, they are in fact difficult to square with the motivation behind the claims that we do endorse. To my mind, however, the strongest argument for full nonvacuous β-conversion (and thus intensional β-equivalence) is how difficult and unnatural it is to work in any higher-order system without it. To reject it in type theory is akin to rejecting familiar classical principles in predicate logic; the reasons that tell against the latter also tell against the former. As much as in the first-order case, we have reason to doubt that a system in which it is not possible to carry out "sustained ordinary reasoning" (Feferman 1984: 95) perspicuously reflects the logical structure of the world.

2.4 Tenability: Some Distinctions

The main use to which type-theoretical models in this work are put is the study of the model-theoretic behaviour of abstraction principles; in particular, we are interested in the constraints that various abstraction principles place on the size of the universe. For that purpose, however, the full power of the languages on offer represents overkill.

Our primary interest is in determining when a *purely logical* abstraction principle—a principle with a single abstraction operator and no other nonlogical terms—can

be satisfied. We introduce the notion of 'tenability', drawing on Kit Fine's work (2002: 114), in order to characterize the notions that interest us. Working in a second-order extensional framework, Fine needs only a simple notion of tenability; we have to draw finer distinctions.

Let $\mathscr{L}^{\mathsf{Bas}}$ be the language with $\Sigma_L = \mathsf{BasLC}$ and $\Sigma_C = \emptyset$. Say that a language $\mathscr{L}^{\Sigma}$ is *spartan* just in case $\Sigma_L = \mathsf{BasLC}$ and $\Sigma_C = \{\S\}$ where $\S$ a generalized abstraction operator (i.e., of type $(\tau \to \tau_o)$ for $\tau_o \neq t$). We define the *first-order size* of an SEL-model $\mathfrak{M}$ as $|D_e(\mathfrak{M})|$; of a Gallin model $\mathfrak{M}$ as $|D_e(\mathfrak{M})|$; and of an M-model $\mathfrak{M}$ as $|I_e(\mathfrak{M})|$.

Definition 2.4.1 A sentence ϕ of a spartan language $\mathscr{L}^{\Sigma}$ is *weakly tenable on κ over a model class $\mathscr{C}$* for just in case there exists an $\mathscr{L}^{\mathsf{Bas}}$-model $\mathfrak{M} \in \mathscr{C}$ such that the first-order size of $\mathfrak{M}$ is $\kappa$and there exists an $\mathscr{L}^{\Sigma}$-model $\mathfrak{M}'$ extending $\mathfrak{M}$ such that $\mathfrak{M}' \vDash \phi$.

Definition 2.4.2 A sentence ϕ of a spartan language $\mathscr{L}^{\Sigma}$ is *strongly tenable on κ over a model class $\mathscr{C}$* just in case, for every $\mathscr{L}^{\mathsf{Bas}}$-model $\mathfrak{M} \in \mathscr{C}$ with first-order size κ, there exists an $\mathscr{L}^{\Sigma}$-model $\mathfrak{M}'$ extending $\mathfrak{M}$ such that $\mathfrak{M}' \vDash \phi$.

The notion of logical consequence is the one appropriate for the model class; the notion of a model extension is the standard one, with $\mathfrak{M}'$ and $\mathfrak{M}$ sharing a frame and differing only in the addition of an interpretation for $\S$.

Definition 2.4.3 If ϕ is weakly (resp. strongly) tenable on κ over $\mathfrak{SEL}(\mathscr{L}^{\mathsf{Bas}})$, we say that ϕ is weakly (resp. strongly) E-tenable on κ. If ϕ is weakly (resp. strongly) tenable on κ over $\mathfrak{Gal}(\mathscr{L}^{\mathsf{Bas}})$, we say that ϕ is weakly G-tenable on κ (resp. on (κ, λ)). If ϕ is weakly (resp. strongly) tenable on κ over $\mathfrak{Msk}(\mathscr{L}^{\mathsf{Bas}})$, we say that ϕ is weakly (resp. strongly) M-tenable on κ.

In the case of E-models, however, the distinction between weak and strong tenability collapses.

Fact 2.4.4 A sentence ϕ of a spartan language $\mathscr{L}^{\Sigma}$ is weakly E-tenable on κjust in case it is strongly E-tenable.

Proof. This follows trivially from the fact that any two SEL-models for $\mathscr{L}^{\mathsf{Bas}}$ with the same first-order size are isomorphic. ⊣

We thus speak of ϕ as simply *E-tenable*, without qualification.

We also define a non-model-theoretic notion that we shall also use:

Definition 2.4.5 A sentence ϕ of a spartan language $\mathscr{L}^{\Sigma}$ is *tenable on the universe* just in case $\phi^{\dagger}$ (the fixed-domain Tarskification of ϕ) is true on the intended interpretation.

Frequently, however, use of the full hierarchy is unnecessary.

Specifically, for the purposes of §4.1–3, we need only consider the *second-order fragment* of $\mathscr{L}^{\mathsf{Bas}}$. Let $\mathscr{L}^{2}$ be this fragment, including logical constants and variables of types e and $(e \to t)$ only (following, for example, the presentation in Shapiro (1991)); let $\mathscr{L}^{2\S}$ be the result of supplementing $\mathscr{L}^{2}$ with an abstraction operator § of some type $((\underbrace{e,\ldots,e}_{n} \to t) \to t)$ for $n \geq 1$. We use standard models for $\mathscr{L}^{2}$ and $\mathscr{L}^{2\S}$, and for other second-order fragments extending $\mathscr{L}^{2}$, and we use 'satisfiable' in its ordinary sense. When writing definitions of models, we indicate only the first-order domain D_e, since $D_{(e\to t)} = \mathscr{P}(D_e)$. We say that a sentence ϕ of $\mathscr{L}^{2\S}$ is $\mathscr{L}^{2}$*-tenable* on κ when there exists an $\mathscr{L}^{2}$-model $\mathfrak{M}$ with first-order size κ and an $\mathfrak{M}'$ extending $\mathfrak{M}$ such that $\mathfrak{M}' \vDash \phi$. Since $\mathscr{L}^{2}$-models are fragments of SEL-models, we need not distinguish strong and weak tenability.

III. THREE PATHS OF ABSTRACTION

The antireductionist account of cardinals and ordinals that will be developed in Chapter IV relies crucially on abstraction principles. In particular, the account of ordinal numbers that I shall offer involves the claim that the Restricted Isomorphism-Type Abstraction Principle (RITA),

$$(\text{RITA}) \qquad \forall X^2 \forall Y^2 (\Omega X^2 = \Omega Y^2 \leftrightarrow ((\text{Big Fld}(X^2) \wedge \text{Big Fld}(Y^2)) \vee X^2 \cong Y^2)),$$

plays a constitutive role in characterizing what it is to be an ordinal. Similarly, I take Hume's Principle,

$$(\text{HP}) \qquad \forall X \forall Y (\#X = \#Y \leftrightarrow X \approx Y),$$

to play a constitutive role in characterizing what it is to be a cardinal. Here $\cong$ is an operator of type $((e, e \to t), (e, e \to t) \to t)$ expressing isomorphism between the Relation denoted by X^2 and that denoted by Y^2; $\approx$ is an operator of similar type expressing equicardinality; Ω and $\#$ are the order-type and cardinal abstraction operators, respectively (of types $((e, e \to t) \to e)$ and $((e \to t) \to e)$; and Big Fld is a term of type $((e, e \to t) \to t)$ expressing the fact that the field of the Relation to which it applies is equinumerous with the universe. Whether these operators are to be understood as primitive or as abbreviations for formulas using more familiar connectives will be important in §5.2; here, however, we need not take any stance on the question.

HP also plays a fundamental role in the neologicist account of arithmetic, first developed in detail by Crispin Wright in his monograph *Frege's Conception of Numbers as Objects* (1983). In the following decades, the neologicist account of arithmetic has been the object of extensive debate.[1] On the neologicist account, HP possesses a special

[1] For surveys, see, for example, MacBride (2003), Ebert and Rossberg (2016a); see generally the essays in Hale and Wright (2001b), Boolos (1998), Heck (2011, 2012), and the works of Fine (2002) and Burgess (2005).

status: it functions as an *implicit definition* of the cardinality operator #. Since, by what has come to be known as Frege's Theorem, full second-order arithmetic (Z^2) is derivable in a standard second-order proof system from Frege Arithmetic (FA, the system whose only axiom is HP) in conjunction with appropriate definitions, the implicit definition thesis, if tenable, would provide at least a partial vindication of the programme of reducing arithmetic to logic.

More generally, on the neologicist conception, a licit abstraction principle of the form

$$\forall X \forall Y (\S X = \S Y \leftrightarrow \phi(X, Y)), \tag{3.0.1}$$

serves to define implicitly the § operator figuring therein. (For simplicity, in this chapter, I shall not consider polyadic and higher-type abstraction principles.) On the standard neologicist picture, motivated by an idea found in §64 of Frege's *Grundlagen* (Frege 1884), (3.0.1) manifests the 'recarving' of the right-hand side ($\ulcorner \phi(X, Y) \urcorner$) to yield the left-hand side ($\ulcorner \S X = \S Y \urcorner$). $\ulcorner \S X = \S Y \urcorner$ does not differ in content from $\ulcorner \phi(X, Y) \urcorner$; rather, the same content is merely presented under two guises. 'Content' here is a theoretical term: in order for the neologicist picture to work out, the sameness-of-content relation must partition the space of terms in a fashion finer-grained than the sameness-of-denotation relation but coarser-grained than the sameness-of-sense relation.

Although I draw on many technical results discovered by neologicists and their interlocutors, my use of HP and other abstraction principles is not neologicist in spirit. Instead, it is motivated by an alternative picture of the role abstraction principles should play in our account of mathematics—*abductivist abstractionist realism* (AAR). In slogan form, AAR about abstraction principles is the view that global considerations of theoretical simplicity, elegance, and explanatory power support the claim that true abstraction principles express laws governing the mathematical universe, analogous to the laws of nature that govern the physical universe. According to AAR, some functors

corresponding to abstraction operators are *élite* in David Lewis's sense (Lewis 1983, 1984): they are metaphysically privileged over less natural, gerrymandered competitors, and the role of abstraction principles in our global theory of the mathematical realm reflects this fact.

In §3.1, I set out the general framework of AAR. Here I argue only for a conditional claim: if antireductionism about ordinals and cardinals is true, and abstraction principles provide explanatorily powerful ways of synthesizing the informal theories $\hat{T}_{Ord}$ and $\hat{T}_{Card}$ discussed in §1.2, then we have good reason to believe that cardinal-abstraction and ordinal-abstraction functors are metaphysically privileged. In Chapter IV, I argue for both of the antecedents of the conditional.

In §3.2, I offer a critique of logicism. My focus is not on the best-known objections to neologicism—the Julius Caesar problem and the bad company objection—but on the coherence of the notion of content recarving. I argue that the neologicist must accept some version of the content recarving thesis, for it is the only plausible account that can give a principle such as (3.0.1) the status she requires of it. But, in order for the content-recarving thesis to work, the neologicist must offer a principled account of content as a property of expressions intermediate between sense and reference. No well-motivated account of this type is possible: instead, given neologicist assumptions, content collapses into sense, as can be demonstrated by considering the behaviour of expressions such as $\ulcorner \S X = \S Y \urcorner$ under embedding.

AAR is also to be distinguished from a third account of the role that abstraction principles can play in the philosophy of mathematics—the *trivialist* account of abstraction endorsed by Agustín Rayo (2013, 2016)—which I discuss in §3.3. Rayo's trivialist abstractionism bears some resemblance to neologicism: it can be thought of as the result of reformulating the neologicist analysis of (3.0.1) under conditions of semantic descent, using a higher-order identity operator of the type discussed at the end of §2.3. The trivialist endorses not (3.0.1) (where abstraction-operator-containing and abstraction-operator-free expressions are *mentioned*, not used) but the direct object-

language claim

(3.0.2) $$(\lambda XY\,.\,\S X = \S Y) \equiv (\lambda XY\,.\,\phi(X, Y)).$$

(Recall that $\equiv$ here is a variant on $=_{((e\to t),(e\to t)\to t)}$.) This metaphysical picture blocks the collapse of content into sense faced by the neologicist, but it must confront a different problem. Extending arguments due to Kit Fine (2002: 39–41), I argue that no reasonable general justification for identities of the form (3.0.2) avoids overgeneration. If the deflationist retreats to *ad hoc* principles, however, he will have to accept a special class of abstraction functions—those corresponding to operators eligible to figure in true identities akin to (3.0.2). But accepting metaphysically élite functions eliminates whatever advantage deflationism might have in theoretical economy over AAP; furthermore, since the deflationist is committed to an account of the semantics of expressions with abstraction functors that does considerable violence to surface form, his overall position is significantly less appealing than that of the abductivist realist.

3.1 Abductivist Abstractionist Realism: Characterizing the Position

As noted above, the key idea behind AAR is that true abstraction principles are akin to laws of nature for the mathematical realm. But how should this thesis be understood? A first approximation is suggested by the popular idea (Ellis 2001, Bird 2007) that at least some laws of nature characterize the essential properties of kinds. For instance, an essentialist about physical laws might claim that

(3.1.1) It is true in virtue of the nature of being an electron that every electron is negatively charged.

I shall not offer a defence of physical-law essentialism here. It suffices for my purposes to note that, even if (3.1.1) is implausible as a thesis about electrons, it is implausible for reasons that do not apply to corresponding claims about mathematicalia. Resistance to claims such as (3.1.1) is based on intuitions of modal recombination: there is

some intuitive appeal to the claim that there is a possible world in which the laws of nature are other than they are at the actual world and electrons are positively charged. (Perhaps there is even a world w which differs from ours only in that electrons play the positron role and vice versa, and every spacetime region occupied by an electron in the actual world is occupied by a positron at w and vice versa.) But we have no analogous recombination intuitions about mathematicalia: at least for the non-structuralist, there is no intuitive pressure to hold that there exists a world w where π plays the $\sqrt{2}$ role and vice versa.

Thus, whether or not we think that at least some laws of nature are essence-characterizing for the physical objects to which they apply, we should think that at least sometimes the claim expressed by a true *mathematical* principle $\mathscr{P}$ can be essence-characterizing for the mathematical objects to which it applies. In the case where $\mathscr{P}$ is an abstraction principle, those objects are the abstracts lying in the range of the functor denoted by the abstraction operator figuring in it.

I shall focus on the simplest case, HP. Here, '#' picks out a functor $\mathfrak{f} : \mathscr{D}_{(\overline{e,e\to t})} \to \mathscr{D}_{\overline{e}}$ whose range is Card—the class containing all and only the true cardinals. In contrast to the neologicist, the defender of AAR does not think that we fix a meaning for '#' by laying down HP. Instead, $\mathfrak{f}$ is a 'reference magnet', and much of the work of making it the case that '#' refers to $\mathfrak{f}$ and not to something else is done by the world. For example, let $\mathfrak{f}'$ be some functor which differs from $\mathfrak{f}$ only by a permutation on Card: e.g., whereas $\mathfrak{f}$ has $(\lambda X \,.\, \exists_{=121} x\, X x) \mapsto 121_{card}$ and $(\lambda X \,.\, \exists_{=252} x\, X x) \mapsto 252_{card}$, $\mathfrak{f}'$ has $(\lambda X \,.\, \exists_{=121} x\, X x) \mapsto 252_{card}$ and $(\lambda X \,.\, \exists_{=252} x\, X x) \mapsto 121_{card}$. The proponent of AAR believes that $\mathfrak{f}$, rather than $\mathfrak{f}'$, is the semantic value of '#' partly because it is, by its very nature, more suitable for the task: the mere fact that the equivalence relations associated with $\mathfrak{f}$ and $\mathfrak{f}'$ partition $\mathscr{D}_{(\overline{e,e\to t})}$ into the same equivalence classes does not make $\mathfrak{f}$ and $\mathfrak{f}'$ equally acceptable candidates for the role. He sees the situation as analogous to the relation between the actual semantic value of 'electron', whose extension is given by $(\lambda x \,.\, x$ is an electron$)$, and a competitor whose extension is given by $(\lambda x . x$ is an electron $\vee\, \phi(x))$ where ϕ is a gerrymandered condition that applies only to a single non-electron in a

distant possible world. Irreducible facts about comparative naturalness form part of the reason that the former, rather than the latter, plays the semantic role that it does.

It is a key tenet of AAR that the meaning of an abstraction operator is not wholly parasitic on the abstraction principle in which it figures: the situation is different from one in which we introduce a singular term by definition as $(\imath x \,.\, \phi(x))$ after having already established $\exists!x\, \phi(x)$, or when we introduce a purely theoretical predicate through a Ramsey sentence. Instead, we have enough independent access to ƒ that the truth of HP could, in principle, be a genuinely surprising discovery about the world. (Indeed, as a matter of history, the truth of HP in infinite cases *was* a surprising discovery, but not one made by Hume: it was made by Cantor, although arguably presaged by Galileo in *Due nuove scienze*.[2] That HP holds in finite cases was, of course, not surprising: someone who was ignorant of that would lack a piece of knowledge required for adequate command of ordinary applied arithmetic. But there is no difference in principle between the cases: we can imagine beings who possess the concept of number, and use that concept to latch on to ƒ, but who falsely believe that HP fails even in some finite cases.[3]) HP, although necessarily true, has no claim to be analytic. It is knowable a priori, for it is a mathematical truth, and any mathematical truth is knowable a priori if at all. Nonetheless, the methods by which we can gain this knowledge have little in common with the ones we can use to come to know (for example) that all and only female foxes are vixens.

Furthermore, according to AAR, abstraction principles are tied to explanation. This comes in two varieties: first of all, an abstraction principle provides a *synthesizing explanation*: it may justify a wide variety of derived results in a way that is particularly

[2]See Mancosu (2017) for an extensive historical discussion. Mancosu holds that the pre-Cantorian concept of 'being the same size as' was not fully precise, and that mathematicians in effect *decided* to adopt a standard for equality of numbers on which HP is true.

[3]As a matter of fact, we need not use our imaginations: if we hold that strict finitists and scorched-earth nominalists are fully committed to what they write (on the ordinary understanding), then we must credit them with the belief that HP is false in some finite cases, for they do not believe that there are infinitely many objects. Of course, we always have the option of hostile translation: we can finesse the logical form of their claims so as to make them seem less outrageous to those of orthodox views, in the process attributing to them an implicit belief—such as the truth of HP—that they dare not explicitly avow. On the use of such translations in dealing with dialetheists, see Lewis (1990); for a general critique of the approach, see Williamson (2007: 92–95).

illuminating. For instance, HP plausibly provides a unified characterization of the natural number structure (as the finite cardinals under the natural ordering) in a way in which the second-order Dedekind–Peano axioms, taken piecemeal, do not. This is a *cognitive-priority* explanation, not a metaphysical one (in the sense in which the grounding claims discussed in Chapter V are metaphysical explanations); nonetheless, as will be argued there, if p is a cognitive-priority explanation for q, and all other requirements for p's being a metaphysical explanation for q are met, then this fact should raise our credence that p is *also* a metaphysical explanation for q: we are entitled to assume that our standards for understanding at least sometimes track metaphysical structure.

In addition, however, an abstraction principle may be *internally explanatory*—the asymmetry in its right-hand and left-hand sides can itself track metaphysical structure. The availability of these explanations provides one initial reason for looking to abstraction principles to account for facts about cardinals and ordinals. It is clear that there is some tight relationship between its being the case that the number of Galilean moons of Jupiter is the same as the number of seasons and its being the case that there is a bijection between the Galilean moons of Jupiter and the seasons. HP provides a candidate explanation: the number of Galilean moons of Jupiter is the same as the number of seasons *because* there is a bijection between the Galilean moons of Jupiter and the seasons. (If the natural numbers are the finite von Neumann ordinals, in contrast, then any explanation of the number of Galilean moons of Jupiter's being the same as the number of seasons will pass through facts about $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}$, which is far less elegant.) This kind of explanatory framework again parallels what we find in the physical sciences. For instance, the following physical law supports asymmetrical internal explanation:

(3.1.2) For any path π between spacetime points p and q, a particle travelling from p to q acted upon by gravity alone will follow π if and only if π is a geodesic.

The fact that π is a geodesic is a good explanation of the fact that particle a, acted upon by gravity alone, follows π; the fact that particle a, acted upon by gravity alone, follows π is not a good explanation of the fact that π is a geodesic.

Thus, the basic idea behind AAR ties a systematizing theory of mathematics to explanation and essences: *abstraction principles about cardinals and ordinals are mathematically powerful because they are internally and externally explanatory, and the best explanation for their explanatory power is that they tell us something about the essences of abstraction functors.*

Why should we accept the truth of AAR (conditional on the claims, to be defended later, that abstraction principles are mathematically powerful and that the most plausible alternative account—reductionism—is false)? There are both negative and positive arguments for the conclusion. The negative argument is that no competing account of the role played by abstraction principles in mathematics can do the explanatory work required: it is the purpose of §§3.2–3 to support this claim by showing that neologicism and trivialism are both flawed. AAR escapes the flaws because it avoids the problematic commitments of the other views.

The positive argument for AAR is that it accounts for the explanatory power of abstraction principles in a way that coincides with a plausible picture of how our best theories of the non-mathematical world relate to their objects (the essentialist account of laws of nature). For a realist about mathematics, having such an account is highly desirable: just as our semantics should not treat mathematical language disanalogously to other kinds of language unless required (Benacerraf 1973), our account of mathematical *theory structure*—the ways in which some aspects of a theory have a privileged position and are tied to others by relations of explanatory dependence—should, if possible, parallel the account used outside of mathematics.

We thus have a multi-stage abductive case for AAR: We start from some basic data about mathematics that we systematize, in more or less the manner presented in Chapter I. We show that reductionism is unpromising, and we develop a non-reductionist system using abstraction principles. In the terminology introduced by Gödel, this sys-

tem is supported both by 'intrisic' evidence (broadly speaking, the internal coherence and theoretical elegance of the framework) and 'extrinsic' evidence ("fruitfulness in consequences" (Gödel 1964: 261), the fact that it recovers the basic data appropriately within a powerful system). Of these two, extrinsic evidence plays the larger role. But this is typical of inquiry in the foundations of mathematics; long before Gödel, Russell emphasized the role played by this kind of abductive fitting of principles to known results in his arguments for adding the Axiom of Reducibility to the type system of *Principia*:

> That the axiom of reducibility is self-evident is a proposition which can hardly be maintained. But in fact self-evidence is never more than a part of the reason for accepting an axiom, and is never indispensible. The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it. (Whitehead and Russell 1927: 1.59).

In the final step, we argue that AAR is the *metaphysical* theory that provides the best explanation for the *mathematical* success of abstractionist nonreductionism.

The overall acceptability of this procedure, of course, depends on the more general question of the acceptability of abductive methods in philosophy and the foundations of mathematics. This is not the place for a full defence of inference to the best explanation. The most persuasive argument for it is that offered by Timothy Williamson (2016, 2017) on the basis of general anti-exceptionalism concerning philosophical methodology: abductive reasoning is pervasive in other forms of human inquiry; if we accept it as reliable there (eschewing corrosive scepticism) and we reject the absence of a firm demarcation line between philosophy and mathematics and other sciences (following in the broadly naturalistic tradition inaugurated by Quine), then we should also accept

it as reliable when applied to the sort of questions under consideration here.

3.2 Neologicism and Content Recarving

In contrast to the abductivist realist, the neologicist believes that we give an abstraction operator its meaning by laying down an abstraction principle: an abstraction principle *implicitly defines* the operator that occurs in it. Wright and Bob Hale describe the view thus:

> An explicit definition aims to supply a semantically equivalent expression of the same syntactic type as the definiendum. Implicit definition, taken as the complement of explicit, embraces a variety of subtypes. What all have in common is the idea that we may fix the meaning of an expression by imposing some form of constraint on the use of longer expressions—typically, whole sentences—containing it. On one traditionally influential conception, this constraint is imposed by the (putatively) *free stipulation* of the truth of the certain sentence, or range of sentences, embedding the *definiendum* and composed of otherwise previously understood vocabulary. [...] We shall refer to the thesis that at least some kinds of non-inferential a priori knowledge are founded in implicit definition as *the traditional conception*. (Hale and Wright 2000: 117[4])

According to the neologicist, if the abstraction principle meets whatever constraints are required for a good implicit definition, then, simply in virtue of its being an implicit definition, one can use it to gain knowledge about the abstracts that lie in the range of the function corresponding to the abstraction operator—moreover, this knowledge will have a special status. The status is not mere apriority: Gödelian intuition, appeals to background metaphysical knowledge, the application of abductive criteria, and the use of thought experiments are all ways of gaining a priori knowledge, but the neologicist will not be satisfied with any account of knowledge about mathemat-

[4]Footnotes omitted; cf. Wright (1983: 104) and Hale and Wright (2001b: 11–14).

ical abstracts on which these procedures have an ineliminable role. Instead, what is desired is knowledge "without further epistemological presupposition" (Wright 1997: 276)—knowledge that is free from reliance on anything but logic, background linguistic knowledge, and reflection on the content of the implicit definition itself. We can call this kind of knowledge *pristine* knowledge. The neologicist programme about abstraction then factors into two theses:

(3.2.1) At least some substantive abstraction principles are licit implicit definitions of the abstraction operators figuring in them.

(3.2.2) If an abstraction principle $\mathscr{A}$ is a licit implicit definition of the abstraction operator figuring in it, then by laying down $\mathscr{A}$ one is in a position to obtain pristine knowledge about the abstracts lying in the range of the function denoted by the abstraction operator therein.

In the simplest case, the knowledge involved in (3.2.2) is simply the knowledge of the proposition expressed by $\mathscr{A}$ itself. But plausibly the neologicist should also maintain that some forms of supplemental knowledge are available beyond this: for instance, on various responses to the Julius Caesar objection advanced by Hale and Wright (2001a), we can, through the process of laying down an abstraction principle, come to gain pristine knowledge of *sortal facts* about the abstracts in question (*e.g.*, that the number of protons in a cesium atom is not a person). Less controversially, the neologicist will presumably also say that we can come to gain pristine knowledge of the consequences of the *reflective closure* of a good abstraction principle. For instance: the result of adding HP to second-order logic (with full comprehension) is FA, the system of Frege Arithmetic. If neologicism is correct, an agent who lays down HP should be able to reflect upon the fact that every logical axiom of second-order logic and every proper axiom of FA is true in the intended model, and that every proof rule of the standard proof system for second-order logic preserves truth. She should thus be in a position to come to know the local reflection principle for FA, $Bew_{\mathsf{FA}}\ulcorner\phi\urcorner \rightarrow \phi$, and thus that FA is consistent—which, of course, exceeds what she can derive directly from FA itself.

The crucial importance of (3.2.1) and (3.2.2), from the point of view of neologicists such as Hale and Wright, lies in a thesis that connects the original logicist programme with the modern critiques, due to Field and Benacerraf, of various purported mechanisms of a priori mathematical knowledge:

(3.2.3) If a priori mathematical knowledge is possible at all, then pristine mathematical knowledge is possible.

In contrast, the defender of AAR, or some other non-neologicist approach to understanding abstraction principles, has no reason to accept (3.2.3): she does not begin from a position of scepticism about non-pristine ways of coming to know about mathematical objects.

Even if (3.2.3) is granted, however, the neologicist faces the task of making out a plausible notion of 'implicit definition' on which (3.2.1) and (3.2.2) are both true. The content recarving thesis is an attempt to do this. The inspiration for the content recarving thesis is to be found in *Grundlagen* §64:

> The judgement "line a is parallel to line b", or, using symbols,
>
> $$a \parallel b,$$
>
> 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 $\parallel$ 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. (Frege 1884: §64/1980: 74)[5]

[5] "Das Urtheil: »die Gerade a ist parallel der Gerade b«, in Zeichen: $a \parallel b$, kann als Gleichung aufgefasst werden. Wenn wir das thun, erhalten wir den Begriff der Richtung und sagen: »die Richtung der Gerade a ist gleich der Richtung der Gerade b«. Wir ersetzen auch das Zeichen $\parallel$ durch das allgemeinere $=$, indem wir den besonderen Inhalt des ersten an a und b vertheilen. Wir zerspalten den Inhalt in anderer als der ursprünglichen Weise und gewinnen dadurch einen neuen Begriff" (trans. Austin).

Clearly, it will not do to identify content with sense, if anything akin to a cognitive significance criterion of sameness of sense is to be preserved, for the neologicist clearly wishes to allow for cognitive accomplishment in the application of abstraction principles—even if this accomplishment must be tied sufficiently closely to the definitional character of HP for the results of the accomplishment to remain pristine. Nor can the neologicist identify content with denotation, if she follows Frege and maintains that $\overline{\mathscr{D}}_t$ contains only two propositions (the True and the False): this is far too blunt to do the work required. Instead, she must claim that content is something coarser-grained, intermediate between sense and denotation. Hale (1997: 97) describes this as a "weaker notion of sense": on this picture, narrow analyticity—the status possessed by ordinary *explicit* definitions—stands to sense as the *broad analyticity* of HP, taken as an implicit definition, stands to content.[6] Although broadly analytic truths are cognitively non-trivial in a way that narrowly analytic truths are not, they are nonetheless knowable pristinely, without invoking any 'extrinsic' epistemic resources.

On this basis, we can reconstruct the content-recarving argument for (3.2.1), as applied to HP, as follows:

(3.2.4) The content of '$\#X = \#Y$' = the content of '$X \approx Y$'.

(3.2.5) If the content of a formula ϕ with exactly variables $\vec{v}$ free = the content of a formula ψ with exactly variables $\vec{v}$ free, then $\ulcorner \forall \vec{v}(\phi \leftrightarrow \psi) \urcorner$ is a licit implicit definition of any previously unintroduced operator occurring therein.

(3.2.6) HP is a licit implicit definition of the # operator.

The argument carries over to other abstraction principles. I state the argument here using open formulas, in keeping with the neologicist tradition. There are a number of ways that this can be understood, but the simplest is to take the content of an open formula to be derived from the more basic notion of *content relative to an assignment*:

[6] See also Yablo (2008) on this point.

we can assume that, where ϕ and ψ are formulas with exactly $\vec{v}$ free, the content of ϕ is identical to the content of ψ if, for every assignment $\mathfrak{a}$ of values to $\vec{v}$, the content of ϕ relative to $\mathfrak{a}$ is identical to the content of ψ relative to $\mathfrak{a}$. Arguably, since quantification in the type system we use is understood through a combination of quantificational constants and λ-binding, it is more natural to speak of the content of '$(\lambda X \lambda Y \,.\, \#X = \#Y)$' and the content of '$(\lambda X \lambda Y \,.\, X \approx Y)$'. In §3.3, I shall use something akin to this approach; here, however, the distinction is immaterial, so I follow convention.

The difficulty lies in justifying (3.2.4). Attempts to give a full *analysis* of the sameness of content relation, such as Hale's (1997) notion of "compact entailment", have fared badly.[7] The demand for an analysis, however, is too stringent: the neologicist, like others invoking a novel piece of conceptual apparatus, is *prima facie* justified in proceeding in a less formal fashion, providing some general principles and relying on paradigm cases to get her meaning across. This *prima facie* justification, however, does not relieve her of the burden of responding to direct challenges to her claims: she need not provide an exhaustive theory to back up (3.2.4), but she does need to show that it is not the case that any theory licensing (3.2.4) will be incoherent. The most serious challenge is the *argument from embedded contexts*, to which, I maintain, the neologicist has no good response.

The argument from embedded contexts arises from the Fregean picture that the neologicist adopts: terms with the same denotation are intersubstitutable *salva veritate* in ordinary transparent contexts; terms with the same sense are additionally intersubstitutible in oblique contexts which shift the denotation of an embedded term from its customary denotation to its customary sense. Because (according to the neologicist) content lies between sense and denotation in coarseness of grain, we should expect to see intersubstitutivity *salva veritate* of content-sharing expressions over a greater range of contexts than terms which merely share a denotation and a smaller range than those which share a sense. Let us call these contexts, which encompass canonical transparent

[7] Potter and Smiley (2001, 2002); see also Hale (2001). I shall not review the long series of modifications to Hale's proposal, since they all face serious counterexamples.

contexts and at least some others, *translucent*. Hence the relevant intersubstitutivity principle is:

(3.2.7) If ⌜—…—⌝ is translucent with respect to the '…' position and the content of a formula ϕ with exactly variables $\vec{v}$ free = the content of a formula ψ with exactly variables $\vec{v}$ free, then ⌜—ϕ—⌝ is true if and only if ⌜—ψ—⌝ is.

Thus, if there are translucent contexts in which '$\#X = \#Y$' and '$X \approx Y$' are not intersubstitutible *salva veritate*, (3.2.4) fails.

But consider the operator 'it is a model-theoretic logical truth that' (i.e., an object-language operator corresponding to our $\models_E$ in Definition 2.2.4 or to $\models$ in some similar system of higher-order logic). We need not assume that the neologicist takes this operator to represent the *correct* account of logical truth; in fact, it is far more natural for her to understand 'it is a logical truth that' as a primitive operator, corresponding to a non-model-theoretic interpretation of her language. All that is required is that she accept it as a *coherent* operator that she can introduce. If it generates translucent contexts, then we can argue as follows:

(3.2.8) ⌜It is a model-theoretic logical truth that $\exists X \exists Y \neg(\ldots) \rightarrow \exists x \exists y\, x \neq y$⌝ is translucent with respect to the '…' position,

(3.2.9) 'It is a model-theoretic logical truth that $\exists X \exists Y \neg(\#X = \#Y) \rightarrow \exists x \exists y\, x \neq y$' is true,

(3.2.10) 'It is a model-theoretic logical truth that $\exists X \exists Y \neg(X \approx Y) \rightarrow \exists x \exists y\, x \neq y$' is false

(since '$\exists X \exists Y \neg(X \approx Y)$' is satisfied in a model with one object in the first-order domain), yielding the result that the content of '$\#X = \#Y$' is not identical to that of '$X \approx Y$'.

But is ⌜It is a model-theoretic logical truth that $\exists X \exists Y \neg(\ldots) \rightarrow \exists x \exists y\, x \neq y$⌝ translucent? Clearly, if it fails to be translucent, the source of this failure must come from

the 'it is a model-theoretic logical truth that' operator. This is not straightforwardly like a belief or knowledge operator or some other context in which, on the Fregean picture, a term denotes its customary sense. Instead, it closely resembles a modal operator (indeed, on some accounts of 'logical necessity', it *is* a modal operator), and it certainly appears that the neologicist should accept that at least some modal operators are translucent. (This is independent of the question whether the operator is *worldly*, in the sense discussed in §2.2.) It is no objection to the translucency of this context that, although it is a model-theoretic logical truth that Hesperus = Hesperus, it is not a model-theoretic logical truth that Hesperus = Phosphorus. Since we have independent reasons for thinking that 'Hesperus' and 'Phosphorus' will differ in content (for they differ in sense and have no internal structure to recarve), this is exactly what we should expect. Thus the burden lies with the neologicist to give reasons why 'it is a model-theoretic logical truth that' does not generate translucent contexts.

There are two main objections to this argument. First, it might be claimed that 'it is a model-theoretic logical truth' is straightforwardly non-translucent because it is *quotational*, and even synonyms cannot be substituted *salva veritate* in quotational contexts. The idea behind this objection is that the status of the operator as part of the object language is factitious. In reality, the objector claims, it is merely a disguise for the metalinguistic $\models$, which attaches not to a sentence but to the *name* of a sentence.[8] (After all, this was how we motivated the operator in stating the argument.)

But the heuristic motivation does not suffice on its own to establish quotational status: we frequently move back and forth between use and mention in explaining operators. What is it about the truth-conditions for 'it is a model-theoretic logical truth that' that could render the expression quotational? We can state the truth condition as follows:

[8] See Burgess (1998: 169) on 'endometalogic' and Scott (1971) on related issues. A kindred objection (suggested by Timothy Williamson) is that the operator is unacceptable because it is a 'monster' in the sense of Kaplan (1989: 510)—an operator that acts on character rather than content. But Kaplan's objections to monsters have to do with the fact that embedded indexicals in natural languages generally seem to take primary scope; if desired, standard syntactic resources for differentiating scope at the level of logical from that at the level of surface form can preserve this feature without affecting the behaviour of 'it is a variable-domain logical truth that' in the indexical-free contexts that interest us.

(3.2.11) $[\![$It is a model-theoretic logical truth that $\phi]\!]_{\mathfrak{M}} = 1$ if and only if, for all models $\mathfrak{M}'$, $[\![\phi]\!]_{\mathfrak{M}'} = 1$.

(Here we leave the details of the restriction to a model class tacit.) Again, the mere fact that ϕ is mentioned here does not suffice to make the operator quotational; in any statement of truth-conditions in a metalanguage, expressions from the object language will be mentioned. Nor does it seem that quantification over models, rather than reference to a single model, should suffice to render the operator quotational: where ϕ contains no other instances of 'it is a model-theoretic logical truth that', the value of ϕ on each model $\mathfrak{M}'$ will be determined by an unobjectionable, perfectly standard recursion clause, and any sequence of nested uses of the operator must terminate in such a circumstance.

Moreover, there is a similar operator in modal logic, discussed by Crossley and Humberstone (1977: 18), which is not normally treated as quotational: 'fixedly', whose truth-conditions relative to a world w and a model $\mathfrak{M}$ are:

(3.2.12) $[\![$Fixedly $\phi]\!]^{w}_{\mathfrak{M}} = 1$ if and only if, for all $\mathfrak{M}'$ similar to $\mathfrak{M}$, $[\![\phi]\!]^{w}_{\mathfrak{M}'} = 1$,

where a pointed model (i.e., a model equipped with a designated actual world) $\langle W', w'_{o}, R', v' \rangle$ is similar to $\langle W, w_{o}, R, v \rangle$ just in case $W' = W$, $R' = R$, $v' = v$, and $w'_{o} \in W$.

'Fixedly' is, in effect, merely the projection into a less rich framework of the *diagonal proposition* in two-dimensional modal logic (Chalmers 2006). In fact, both (3.2.11) and (3.2.12) can be treated as special cases of truth conditions for a *family* of operators:

(3.2.13) $[\![\blacksquare_{R}\phi]\!]^{i}_{\mathfrak{M}} = 1$ if and only if, for all $\mathfrak{M}'$ bearing R to $\mathfrak{M}$, $[\![\phi]\!]^{\pi_{\mathfrak{M}'}(i)}_{\mathfrak{M}'} = 1$,

where i is a collection of indices, R is an equivalence relation, and $\pi_{\mathfrak{M}'}$ is a (possibly trivial) mapping of the indices appropriate to $\mathfrak{M}$ to those appropriate to $\mathfrak{M}'$. It seems that either all operators falling under (3.2.13) are quotational or none are. But the former conclusion seems absurd: if we define R for an operator $\boxtimes$ to be the relation a pointed model $\mathfrak{M}' = \langle W', w'_{o}, R', v' \rangle$ bears to $\mathfrak{M} = \langle W, w_{o}, R, v \rangle$ just in case $\mathfrak{M}'$ results

from $\mathfrak{M}$ solely by permuting the world of evaluation with some (not necessarily distinct) $w \in W$. But this results in nothing more than the broadest S5 necessity operator: $[\![\boxtimes\phi]\!]^{w}_{\mathfrak{M}} = 1$ if and only if, for all $w^* \in W$, $[\![\boxtimes\phi]\!]^{w^*}_{\mathfrak{M}'} = 1$. And that operator is surely not quotational.

The second objection that might be mounted against the argument in (3.2.8–10) is that it proves too much. It seems that if 'it is a model-theoretic logical truth' is translucent, then 'it is an intuitionistic logical truth' should be as well. But in that case an argument parallel to (3.2.8–10) would prove that the content of 'Hannibal crossed the Alps' differs from that of '¬¬Hannibal crossed the Alps'; on such a fine-grained picture, content would differ little from sense. There are two replies to this objection. First, it is not clear that 'it is an intuitionistic logical truth' is on a par with 'it is a model-theoretic logical truth'. The latter corresponds to a notion that can be defined precisely in set-theoretical terms (whether or not it is an adequate explication of our notion of logicality). But if classical logic is the true logic, then there is no semantics for 'it is an intuitionistic logical truth' that faithfully mirrors the *intended* interpretation, for the intended interpretation relies on a false picture of the world. We can of course still construct Kripke models for intuitionistic logic, but these will not genuinely illuminate features of the correct logic. Instead, they will be artefacts designed with a particular *proof procedure* in mind: they will capture facts about which sentences can be derived from other sentences using the restricted tools of intuitionistic logic. And this is precisely what we should expect based on the BHK semantics for intuitionistic logic, where the central role is played by the presence or absence of canonical proofs. Yet a canonical proof of a sentence from another sentence is a quotational notion if anything is: it belongs to syntax, not semantics.

Moreover, even if we grant that a parallel argument can be carried out with 'it is an intuitionistic logical truth', this does not affect the force of (3.2.8–10) against the neologicist. To believe otherwise is to misunderstand the dialectical situation: it is only the *neologicist* who needs a notion of content coarser-grained than sense that nonetheless validates (3.2.7); her opponent has no independent commitment to the thesis that such

a notion can be made out. If identity of any sort of content that validates (3.2.7) reduces to identity of sense, then broad analyticity collapses into narrow analyticity—but this is not an unwelcome result for anyone but the neologicist. In fact, it is precisely what one sceptical of the appeal to content, as an independent level between sense and denotation, should expect.

3.3 Trivialist Abstractionism

In contrast to neologicism, the trivialist version of abstractionism defended by Rayo is first and foremost a metaphysical view: what is characteristic of trivialism about HP is acceptance of a 'just is' statement:

> Just like one can distinguish between trivialist and non-trivialist varieties of mathematical platonism, one can distinguish between trivialist and non-trivialist varieties of neo-Fregeanism. Both versions of neo-Fregeanism agree that numbers—the referents of numerical-terms—constitute a realm of mind-independent objects. But they disagree about whether the existence of this realm of objects is a trivial affair.
>
> Trivialist neo-Fregeans go beyond mere acceptance of Hume's Principle; they also accept the 'just is'-statement corresponding to Hume's Principle:
>
> $\#_x(F(x)) = \#_x(G(x)) \equiv_{F,G} F(x) \approx G(x)$
>
> (*Read*: for the number of *F*s to equal the number of *G*s *just is* for the *F*s to be in one–one correspondence with the *G*s)
>
> A consequence of this 'just is'-statement is that the assumption that there be no numbers is not just false, but *absurd*. (Rayo 2013: 77)

Some exegetical difficulty arises because of Rayo's metametaphysical views, but on the most natural construal, 'just is' claims are claims about a way the world is: "one can think of logical space as determined by the set of true 'just is'-statements" (2013: 50).[9]

[9]Rayo denies that there is a unique, non-conceptual-scheme-relative objectively correct way the world is, but he makes clear that nothing in his first-order views about the 'just is' operator turns on this (2013:

Rayo explains that "it is natural to think of '≡' [his formalization of the 'just is' operator] as an *identity predicate* of sorts" (2013: 66; cf. Rayo 2016: 204). Rayo's additional commitments about the 'just is' locution—that it expresses sameness of truth-conditions and that it is equivalent to necessary coextensiveness (for an appropriate necessity operator) (2013: 48–56, 2016: 205–8)—are not required for the application of his account to mathematics. The similarity to Dorr's 'to be an F is to be a G' locution (2016) is striking, and I contend that the best interpretation of Rayo's 'just is' operator is as literal higher-order identity (our $=_\tau$ operator). The use of the quantified notation $\equiv_{\vec{v}}$ flanked by open formulas, where $\vec{v}$ is a series of variables, is also common to Dorr (2016: 46–48); following the latter's lead, we shall understand it as an abbreviation for identity of λ-terms, in keeping with our general approach to variable binding.

Thus Rayo's central claim, as applied to HP, is

$$(\lambda XY \,.\, \#X = \#Y) \equiv (\lambda XY \,.\, X \approx Y). \tag{3.3.1}$$

(3.3.1) is incompatible with AAR. The proponent of AAR maintains that HP is internally explanatory, in the sense that the truth of the right-hand side of (a given instance of) HP *accounts for* the truth of the left-hand side, and that this explanation tracks a real metaphysical difference. But if the higher-order Relation corresponding to the condition that the number of things falling under a Concept X is identical to the number of things falling under a Concept Y simply *is* the higher-order Relation of the bijectibility of the Concept X and the Concept Y, then there is no prospect of a genuine asymmetry that can do explanatory work.

Is (3.3.1) plausible? Rayo's positive arguments for the view depend on a picture he terms 'anti-metaphysicalism', which eliminates the need to answer certain questions by denying the deep coherence of the (seemingly epistemically, if not metaphysically, possible) scenarios that motivate the questions: "By accepting the claim that for there to be a table *just is* for there to be some things arranged tablewise, one eliminates the

63).

need to address an awkward question: what would it take for a region that is occupied by some things arranged tablewise to also be occupied by a table?" (2013: 22). A philosopher who views metaphysical inquiry as continuous with other investigations about the world, and wants to give the terms used in metaphysical theorizing a semantics that coheres fully with the one given to other terms elsewhere, will be less sympathetic to this goal.

In any case, however, there are other direct arguments against (3.3.1). Kit Fine, in an argument directed against the neologicist account of content-sharing (which he sees as a form of synonymy), develops two puzzles based on chains of counterintuitive content-equivalence (2002: 39–41). The puzzles are more compelling, however, when converted into an object-language argument against a view like Rayo's, since in their original form they rely on intuitions about *nonidentity* of content among open sentences. But 'content' is a sufficiently *recherché* notion that it is not clear that a systematic theory is on offer; at least, the neologicist might plausibly contend that the burden is not on her to provide one.[10] (Arguments directed against the neologicist's specific positive claims of identity of content, such as the one in §3.2, do not have this limitation.) Higher-order identity, in contrast, is precisely the sort of thing for which very general systematic principles are desired.

In this framework, Fine's first puzzle can be reformulated as follows. We assume Basic Law V,

$$\forall X \forall Y (\backepsilon X = \backepsilon Y \leftrightarrow \forall x (Xx = Yx)), \tag{BLV}$$

"its inconsistency being irrelevant for present purposes" (Fine 2002: 39). The trivialist accepts not only BLV but

$$(\lambda XY \,.\, \backepsilon X = \backepsilon Y) \equiv (\lambda XY \,.\, \forall x (Xx = Yx)). \tag{3.3.2}$$

[10] See Cook and Ebert (2004: 793–96) for other objections to the original argument.

Following Fine, we argue thence to an unacceptable conclusion as follows:

$$(\lambda XY\,.\,\in X = \in Y) \equiv (\lambda XY\,.\,\forall x(x = \in X \leftrightarrow x = \in Y)), \tag{3.3.3}$$

$$(\lambda XY\,.\,\forall x(x = \in X \leftrightarrow x = \in Y)) \equiv (\lambda XY\,.\,\in(x = \in X) = \in(x = \in Y)) \tag{3.3.4}$$

$$(\lambda XY\,.\,\in X = \in Y) \equiv (\lambda XY\,.\,\in(x = \in X) = \in(x = \in Y)). \tag{3.3.5}$$

(3.3.4) is an instance of (3.3.2), and (3.3.5) follows from (3.3.4) and (3.3.3) trivially. The idea is that (3.3.3) should be uncontroversial: "Indeed, it is hard to see how a criterion of synonymy could be so tolerant as to allow the synonymy of ['$\forall x(Fx \leftrightarrow Gx)$'] to ['$\in X = \in Y$'] (or of ['$(\lambda XY\,.\,\in(x = \in X) = \in(x = \in Y))$'] to ['$(\lambda XY\,.\,\forall x(x = \in X \leftrightarrow x = \in Y))$']) and yet not tolerant enough to forbid [*recte*: allow?] the synonymy of ['$(\lambda XY\,.\,\forall x(x = \in X \leftrightarrow x = \in Y))$'] to ['$\in X = \in Y$']" (Fine 2002: 40).[11]

It is intuitively difficult to deny (3.3.3) in a context where BLV is in play. But it might be objected that Fine's claim that BLV's inconsistency is "irrelevant for present purposes" is too sanguine; it is a familar occurrence that inconsistencies infect seemingly straightforward arguments in problematic ways, and there is no guarantee that we can transfer these claims about $\in$ to other, better-behaved, abstraction operators. But this response is not entirely convincing here, since the argument from BLV and (3.3.2) to (3.3.5) can be carried out in a setting with only Δ^1_1-comprehension, in which BLV is consistent (Ferreira and Wehmeier 2002).[12]

In any case, Fine (2002: 40) has a different argument that can be used directly against (3.3.1). Let X^+ denote the least natural number that is non-X. We proceed as follows:

$$(\lambda XY\,.\,X \approx Y) \equiv (\lambda XY\,.\,(\lambda x.Xx \vee x = X^+) \approx (\lambda x.Yx \vee x = Y^+)), \tag{3.3.6}$$

[11] I have spelt out Fine's labels, as translated into λ-notation. I assume that 'forbid' must be a drafting error, as it renders the sentence nonsensical.

[12] I owe this point to Volker Halbach.

(3.3.7) $(\lambda XY \,.\, (\lambda x.Xx \vee x = X^{+}) \approx (\lambda x.Yx \vee x = Y^{+})) \equiv$
$(\lambda XY \,.\, \#(\lambda x.Xx \vee x = X^{+}) = \#(\lambda x.Yx \vee x = Y^{+}))$,

(3.3.8) $(\lambda XY \,.\, \#X = \#Y) \equiv (\lambda XY \,.\, \#(\lambda x.Xx \vee x = X^{+}) = \#(\lambda x.Yx \vee x = Y^{+}))$.

Again, (3.3.6) is the key premiss, and it is supposed to be intuitively clear. (3.3.7) is an instance of (3.3.1), and (3.3.8) follows from (3.3.1), (3.3.6), and (3.3.7). Now, (3.3.6) is far from as intuitively appealing as (3.3.3), and the Rayo-style trivialist can resist it. But has he any *uniform*, non-*ad hoc* way of doing so? The obvious way is to endorse a compositionality principle for identities of the form

(3.3.9) $(\lambda XY \,.\, X_1 \approx Y_1) \equiv (\lambda XY \,.\, X_2 \approx Y_2)$ only if $X_1 \equiv X_2$ and $Y_1 \equiv Y_2$.

This is satisfiable on an M-model, although care is required: there are similar-looking principles without type-raising that are inconsistent (Goodman 2017). But what right has the trivialist to a principle like this? (3.3.9) is appealing to one who thinks that internal structure in formulations of equivalence relations tracks distinctions at the level of Concepts: deep form, at the metaphysical level, parallels surface form. But the trivialist is forced to reject this governing idea in a particularly brutal fashion when he endorses (3.3.1), where a Concept expressed with singular terms and an identity symbol is asserted to be higher-order identical to one expressed in an entirely different way.

Even if the trivialist can give a non-*ad hoc* reason for rejecting (3.3.6) in this particular case, Fine's arguments generalize. Prescinding from the specific operators and

open formulas involved, we can reformulate the arguments as follows:

(3.0.2) $$(\lambda XY\,.\,\S X = \S Y) \equiv (\lambda XY\,.\,\phi(X,Y)),$$

(3.3.10) $$(\lambda XY\,.\,\S X = \S Y) \equiv (\lambda XY\,.\,\phi((\lambda x\,.\,\psi(\S Y)),(\lambda x\,.\,\psi(\S Y)))),$$

(3.3.11)
$$(\lambda XY\,.\,\phi((\lambda x\,.\,\psi(\S Y)),(\lambda x\,.\,\psi(\S Y)))) \equiv (\lambda XY\,.\,\S(\lambda x\,.\,\psi(\S X)) = \S(\lambda x\,.\,\psi(\S Y))),$$

(3.3.12) $$(\lambda XY\,.\,\S X = \S Y) \equiv (\lambda XY\,.\,\S(\lambda x\,.\,\psi(\S X)) = \S(\lambda x\,.\,\psi(\S Y))),$$

corresponding to (3.3.2–5), and

(3.0.2) $$(\lambda XY\,.\,\S X = \S Y) \equiv (\lambda XY\,.\,\phi(X,Y)),$$

(3.3.13) $$(\lambda XY\,.\,\phi(X,Y)) \equiv (\lambda XY\,.\,\phi(\zeta(X),\zeta(Y))),$$

(3.3.14) $$(\lambda XY\,.\,\S\zeta(X) = \S\zeta(X)) \equiv (\lambda XY\,.\,\phi(\zeta(X),\zeta(Y))),$$

(3.3.15) $$(\lambda XY\,.\,\S X = \S Y) \equiv (\lambda XY\,.\,\S(\lambda x\,.\,\psi(\S X)) = \S(\lambda x\,.\,\psi(\S Y))),$$

corresponding to (3.3.1) and (3.3.6–8).

The trivialist about an abstraction principle licensing (3.0.2) needs to deny that there exists *any* ϕ or *any* ζ such that (3.3.10) (resp. (3.3.13)) is plausible, but (3.3.12) (resp. (3.3.15)) is unacceptable. But without an appeal to compositionality principles similar to (3.3.9), it is far from clear how this is possible. At least, the burden is on the trivialist to give some systematic account of higher-order identity which yields principles like (3.0.2) but does not overgenerate.

The proponent of AAR, on the other hand, who accepts internal explanation—and thus an asymmetric reading of abstraction principles—has no such problem. Moreover, even if the trivialist succeeds in making out a coherent account of higher-order identities *involving some particular abstraction operator*, he will have to explain why this operator is special in its behaviour, for clearly not every true claim of the form $\forall x^{\tau}\forall y^{\tau}(f^{(\tau\to e)}x = fy \leftrightarrow \phi(x,y))$ can license the trivialist thesis $(\lambda x^{\tau}y^{\tau}\,.\,f^{(\tau\to e)}x =$

$fy) \equiv (\lambda xy \,.\, \phi(x,y))$. If the trivialist's answer to this question is that the abstraction operator in question has as its semantic value a functor that is metaphysically privileged—natural and élite—then he will have to explain why this metaphysical privilege does not translate into an internal asymmetry. He will have taken onboard a controversial commitment of AAR—the Lewisian picture of élite functions—without being able to avail himself of AAR's benefits.

IV DEPENDENCE ARGUMENTS

4.1. Dependence, Ordinals, and Cardinals: The Main Argument

In §1.3, I set out a general objection to Benacerraf-style multiple realizability arguments. For the sake of argument, I granted the premiss that for no candidate $K \rightharpoonup \mathit{PureSet}$ reduction f have we any reason to believe that f is correct. Where $\|K\| = \mathbb{N}$, this premiss is reasonable: at the very least, it is intuitively appealing for the two reductions that interest Benacerraf, $\mathbb{N} = \mathbb{N}_{VN}$ and $\mathbb{N} = \mathbb{N}_Z$.

When we consider *ordinal reductions*, however, the situation changes. $\mathbb{N}_Z$, unlike $\mathbb{N}_{VN}$, lacks an extension to limit ordinals, since there can be no infinitely descending $\in$-chains. Hence there is thus no such thing as a proper class On_Z directly generalizing $\mathbb{N}_Z$, and thus no way to develop a full version of *Zermelo reductionism* ($\mathsf{On} = \mathsf{On}_Z$) that would parallel the von Neumann reductionism of $\mathsf{On} = \mathsf{On}_{VN}$.

We can, however, introduce the *quasi-Zermelo ordinals* On_{QZ} as follows: $\mathrm{o}_{QZ} = \emptyset$, $(\alpha + 1)_{QZ} = \{\alpha_Z\}$, and $\delta_{QZ} = \bigcup_{\alpha<\delta} \alpha_{QZ}$ for limit δ.[1] But quasi-Zermelo reductionism ($\mathsf{On} = \mathsf{On}_{QZ}$) is clearly inferior to von Neumann reductionism according to the ordinary criteria of theory choice—implicity, elegance, uniformity, and so on. For instance, although $\delta_{QZ} = \bigcup_{\alpha<\delta} \alpha_{QZ}$ holds for *limit* δ, we do not have $\beta_{QZ} = \bigcup_{\alpha<\beta} \alpha_{QZ}$ for all β_{QZ}. In general, let us say that a function $f : O \to A$ (for some class A of objects and some downward-closed $O \subseteq \mathsf{On}$) obeys the *collection-at-limits principle* just in case, for every limit $\delta \in O$, $f(\lambda) = \bigcup_{\alpha<\delta} f(\alpha)$. Collection-at-limits for the identity mapping on ordinals appears in Zermelo reductionism as an epicycle that cannot be unified in any natural way with the principles for successor ordinals in On_Z; in contrast, $\beta_{VN} = \bigcup_{\alpha<\beta} \alpha_{VN}$ holds for *all* β. (Were there principled reasons to rule out $\mathsf{On} = \mathsf{On}_{VN}$ that did not generalize to reasons to reject reductionism as such, we might be obliged to reconsider $\mathsf{On} = \mathsf{On}_{QZ}$ as a candidate reduction. Later on, I shall discuss a thesis about ordinals that, if successful, would rule out $\mathsf{On} = \mathsf{On}_{VN}$ but not $\mathsf{On} = \mathsf{On}_{QZ}$; I conclude that it

[1] Recall that, in the absence of subscripts, $\alpha, \beta, \ldots$, range over *true* ordinals and $\kappa, \lambda, \ldots$, over *true* cardinals.

lacks plausibility.)

Moreover, the von Neumann reduction is the *only* candidate pure-set reduction for ordinals on which the following desirable equivalence obtains:

$$\forall\alpha\forall\beta(\alpha < \beta \leftrightarrow \alpha \subsetneq \beta \leftrightarrow \alpha \in \beta). \tag{4.1.1}$$

This is not a *decisive* ground for preferring the von Neumann reduction to all others, but it affords at least *some* reason to think it more likely to be correct than the obvious alternatives, and that suffices to make Benacerrafian arguments for reductionism about ordinals even less persuasive than they are in general.

For cardinals, the case is less clear. Given $\mathsf{Card} = \mathsf{Card}_{VN}$, we have nothing like

$$\forall\kappa\forall\lambda(\kappa < \lambda \leftrightarrow \kappa \subsetneq \lambda \leftrightarrow \kappa \in \lambda), \tag{4.1.2}$$

but it is far from clear that this would be desirable, since any $Card \rightharpoonup PureSet$ reduction on which (4.1.2) held would break the connexion—discussed at greater length below—between a cardinal κ and the cardinal $\mathsf{card}(\kappa)$ measuring κ's cardinality. But, *conditional on the truth of* $\mathsf{On} = \mathsf{On}_{VN}$, $\mathsf{Card} = \mathsf{Card}_{VN}$ is far more elegant than any other $Card \rightharpoonup Ord$ reduction; to the extent that $\mathsf{On} = \mathsf{On}_{VN}$ is superior to other $Ord \rightharpoonup PureSet$ reductions, $\mathsf{Card} = \mathsf{Card}_{VN}$ becomes more appealing: here as well, it appears that Benacerraf's argument is in worse shape than in the general case (albeit less so than with the ordinals).

In order to develop alternative antireductionist arguments, I introduce two new pieces of ideology—the notions of *objectual ontological dependence* and *immediate objectual ontological dependence*. Unlike *propositional dependence* or *grounding* (to be discussed in Chapter V), which can be thought of as the regimentation of a certain metaphysically laden usage of the fine-grained dyadic sentential operator 'because', objectual dependence and immediate objectual dependence require no hyperintensional apparatus to formulate: they are simply type-$(e, e \to t)$ primitive constants, whose seman-

tic values are binary Relations among objects. Dependence *simpliciter* is represented by $\lhd$: we read '$x \lhd y$ ' as 'y depends on x'. Immediate dependence is represented by $\lhd_I$: we write '$x \lhd_I y$' for 'y depends immediately on x'. We term $\{z : z \lhd x\}$ the class of x's *dependees*, $\{z : x \lhd z\}$ that of x's *dependants* (and *mutatis mutandis* for immediate dependees and immediate dependants).

As is customary, I shall appeal to paradigm cases to motivate these notions. Ontological dependence is an irreflexive, asymmetric, transitive relation, whose clearest-cut instance—and the one which will be most important for our purposes—is the dependence of a set on its elements. Øystein Linnebo explains the intuition thus:

> According to the prevailing iterative conception, sets are 'formed from' their elements. The relation between a set and its elements is thus asymmetric, because the elements must be 'available' before the set can be formed, whereas the set need not be, and indeed cannot be, 'available' before its elements are formed. A set thus appears to depend on its elements in a way in which the elements do not depend on the set. (Linnebo 2008: 72)[2]

Thus, the set {Nelson, Villeneuve} depends on Nelson and it depends on Villeneuve, but neither Nelson nor Villeneuve depends on {Nelson, Villeneuve}. The relation in question is non-causal and finer-grained than modal dependence: {Nelson, Villeneuve} exists at every possible world in which the two admirals do, but it *owes its being* to them, whereas they do not owe their being to it. The same principle applies where the objects exist necessarily: $\{\emptyset\}$ depends on $\emptyset$, but not vice versa.

We thus have the following *weak basic datum* about dependence:

(WBD) $$\forall x \forall y (x \in y \rightarrow x \lhd y),$$

Moreover, the dependence in question is immediate, for there is no object x such that {Nelson, Villeneuve} depends on Nelson at least partly in virtue of {Nelson, Villeneuve}'s depending on x and x's depending on Nelson (and likewise for

[2] Linnebo mentions Boolos (1971) and Fine (1995b), *inter alia*, as defenders of the view.

{Nelson, Villeneuve}'s dependence on Villeneuve). We can thus strengthen WBD to the *strong basic datum*:

(SBD) $$\forall x \forall y (x \in y \rightarrow x \lhd_I y).$$

Of course, even WBD is not universally accepted: Luca Incurvati (2012: 79), for instance, is sceptical that there is a natural, meaningful relation of metaphysical dependence; even if there is, he claims, it is unclear that we understand it well enough for it to play a significant role in our understanding of the relationship between a set and its elements.[3] But here the situation is much as with any notion whose coherence is contested: there is no certain way of convincing a determined sceptic, but with enough examples one can get the outlines of the concept across. Other examples are more controversial than those involved in SBD, but they are nonetheless useful in fixing intuitions. It is intuitively plausible that a composite material object depends on its proper parts.[4] Likewise, tropes, if such things exist, presumably depend ontologically on the substances in which they inhere. Social objects—nations, clubs, corporations—depend ontologically on the persons from which which they are constituted. Each of these claims can be denied on various grounds, but collectively they delimit the relevant notion in at least a rough and ready way.[5]

Can we reduce immediate dependence to dependence *simpliciter* or vice versa? It is clear that the rough characterization given above—the absence of an object z such

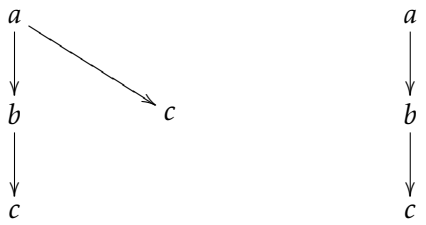
[3] See also Potter (2004: 38–39).

[4] Jonathan Schaffer (2010) has argued against this claim on two grounds: (a) cases of quantum-mechanical entanglement and kindred situations which appear to involve irreducible emergent properties and (b) cases of gunk. With respect to (a), it is unclear that on the best available (field-theoretic) accounts of quantum mechanics, particles (if they exist at all) should be taken to be literal parts of entangled systems; in any case, the thesis that a material object depends on its proper parts *in cases where there are no irreducible emergent properties* is unthreatened. With respect to (b), the existence of gunk is only incompatible with the claim that composite material objects depend on their proper parts if the dependence relation must be well-founded, but there is no reason to think that this is the case.

[5] For other examples, see Koslicki (2012: 188–89). It may be possible to derive SBD from a more fundamental principle. Fine (2010b) offers a general theory of *composition operators*—partial functions from sequences of objects to objects—of which the set-builder operation, the multiset-builder operation, mereological summation, and various others are special cases. It seems that the most discriminating relation on objects induced by each of these composition operators—*immediate parthood*, *immediate multiset-membership*, and so on—obeys an analogue of SBD. The parallel is suggestive but cannot be pursued here.

that $x \lhd y$ at least partly in virtue of $x \lhd z$ and $z \lhd y$—cannot be turned into an explicit, purely extensional analysis: as Fine (1995b: 281–82) notes, it is not possible to determine the graph of the $\lhd_I$ relation over a group of objects on the basis of the graph of $\lhd$ alone. For instance, Structures I and II, with immediate dependence represented by arrows from dependants to dependees, have the same profile in terms of dependence *simpliciter*:

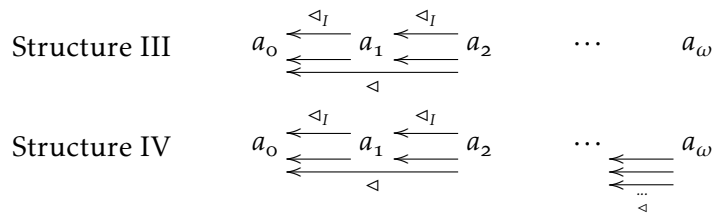
Structure I Structure II



In contrast, Fine argues that "[m]ediate dependence can be defined in terms of immediate dependence; for the one can simply be obtained by chaining the other" (1995b: 281). This suggests that dependence *simpliciter* is the transitive closure of immediate dependence, which we shall write $\lhd_I^*$.

Given reasonable assumptions about the breadth of the metaphysically possible, however, this is incorrect. Consider the case of a homogeneous solid occupying a region of space parameterized along one axis by the real interval $[0, 1]$ (in some specified unit): call this solid A. Let B be the portion of the solid occupying $[0, 0.9]$ along the given axis and C that occupying $(0.9, 1]$. Furthermore, assume that there are no irreducible emergent properties of A to complicate the situation: all the physical facts about A are accounted for (in a suitable sense) by the laws of nature and the physical facts about B and C. It seems intuitively plausible that A should depend objectually, in a mediate fashion, on its proper part B and on its proper part C; yet there is no chain of finitely many intermediate objects, larger than B but smaller than A, connected by relations of immediate dependence that can be used to ground A's dependence on B (and *mutatis mutandis* for A and C).

A more abstract countable example is given by Structures III and IV (of order type $\omega + 1$, with carrier sets $\{a_i : i < \omega + 1\}$):



In Structure III, $a_i \not\lhd_I a_0$ and $a_i \not\lhd a_0$ for all i; for finite j, $a_i \lhd_I a_j$ if and only if $j = i + 1$ and $a_i \lhd a_j$ if and only if $i < j$; and $a_i \not\lhd_I a_\omega$ and $a_i \not\lhd a_\omega$ for all i. In Structure IV, $a_i \not\lhd_I a_0$ and $a_i \not\lhd a_0$ for all i; for finite j, $a_i \lhd_I a_j$ if and only if $j = i + 1$ and $a_i \lhd a_j$ if and only if $i < j$; and $a_i \not\lhd_I a_\omega$ for all i, but $a_i \lhd a_\omega$ if and only if $i < \omega$. If $\lhd$ must coincide with $\lhd_I^*$, then only Structure III is metaphysically possible. But there seems to be no good reason why dependence structures akin to Structure IV cannot exist. In fact, I shall argue below that Structure IV corresponds to a reasonable candidate dependence profile for the ordinals up to and including ω. But if instances of Structure III and instances of Structure IV are both metaphysically possible, then facts about immediate dependence do not suffice to determine uniquely facts about dependence *simpliciter*, precluding any reduction of the latter to the former. The implication $\forall x \forall y (x \lhd_I^* y \rightarrow x \lhd y)$ is trivial, but, *pace* Fine, we should not endorse the converse.

With this apparatus in place, we can set out dependence-based arguments for antireductionism about ordinals and cardinals. We begin with the ordinals.

If $\mathsf{On} = \mathsf{On}_{VN}$, then, by SBD, we have the result

(D1) $$\forall \alpha \forall \beta (\alpha < \beta \rightarrow \alpha \lhd_I \beta).$$

But (D1) is, intuitively, a highly implausible dependence profile for ordinals: it entails that the ordinal 6_{Ord} depends *immediately* on the ordinal 4_{Ord}, not merely mediately (e.g., in virtue of 6_{Ord}'s depending on 5_{Ord} and 5_{Ord}'s depending on 4_{Ord}). Likewise, given (D1), ω depends immediately on every finite ordinal, and $\omega + 1$ depends imme-

diately not only on ω but on every finite ordinal.

We can flesh out the intuition behind (D1)'s lack of appeal further as follows: there seem to be two reasonable candidate characterizations of the functional role of ordinals—characterizations that are in themselves neutral as to whether ordinals are sets—that correspond to dependence profiles incompatible with (D1).

On the *pure order-type conception*, to be an ordinal just is to be the privileged representative of an equivalence class of isomorphic well-orderings—the objectual correlate of the (higher-order) structure that every well-ordering of a given order type has in common. To spell this out, let us temporarily return to the informal notation for structures used in §1.2. On the pure order-type conception, everything fundamental about 2_{Ord} is captured by the fact that it 'represents' every structure $\mathscr{S} = \langle S, \prec_S \rangle$ of the form $\langle \{a,b\}, \{\langle a,a\rangle, \langle a,b\rangle, \langle b,b\rangle\}\rangle$ (where $a \neq b$).

On this conception, ordinals are derivative of the structures that they represent; but if this is the case, then an ordinal α depends on an object x just in case, for *every possible* structure $\mathscr{S}$ such that α represents $\mathscr{S}$, x figures in the carrier class of $\mathscr{S}$. But *no* object figures in the carrier class of *every* $\mathscr{S}$ in the equivalence class of structures represented by a given ordinal: thus no ordinal depends on anything,

(D2) $$\forall \alpha \neg \exists x (x \lhd \alpha),$$

and, *a fortiori*, no ordinal depends on any ordinal:

(D3) $$\forall \alpha \forall \beta (\alpha \not\lhd \beta).$$

I have presented the pure order-type conception on the assumption that informal reference to an *abstract* structure—what is common to all concrete structures in an equivalence class under isomorphism—is, like our use of class talk, not to be taken literally. This assumption is one an *ante rem* structuralist might contest; as we saw in §1.2, *ante rem* structuralism coheres naturally with the position that there exists

an object—an A-structure—$\$(\mathscr{S})$, at least in some cases, corresponding to a higher-order structure $\mathscr{S}$. In §4.4, I shall discuss *ante rem* structuralism at greater length, but it should be noted that even if we accepted the existence of A-structures, this in itself would motivate only the claim (assuming the pure order-type conception) that an ordinal α depends on the A-structure $\$(\mathscr{S})$ instantiated by every $\mathscr{S}$ in the equivalence class of well-ordered relations indexed by α.[6] (D2) would fail, but the principle is compatible with (D3), which suffices for the anti-reductionist arguments below. (An additional claim common to *ante rem* structuralists—termed *objects depend on objects* by Linnebo (2008: 67)—is not compatible with (D3), but, as will be noted in §4.5, it is very implausible.)

On the alternative, *bottom-up conception*, ordinals have an additional role to play: it is essential to an ordinal not only that it represent a well-ordering, but that its specific place in the sequence of ordinals be determined by the fact that it *extends* the ordinals smaller than it. If we take the notion of 'extending' sufficiently seriously, we obtain a conception on which ordinals either literally have an internal structure or stand in relation to something that can go proxy for an internal structure, even as they represent the well-orderings in the relevant equivalent class. We shall term whatever internal structure-like entity it is that accounts for the extending relationship the *skeleton* of the ordinal. We can use directed graphs to stand for skeletons. The ordinal 2_{Ord} has the skeleton

and 3_{Ord} the skeleton

of which the skeleton of 2_{Ord} is something akin to a proper part.

On the bottom-up conception, this relation among skeletons (which we shall call

[6] See also Menzel (1986: 43), who seems to endorse an A-structure-based account of ordinals: "Intuitively, though, what [ordinals] have in common cannot be a set, for what they share is a common *structure*, i.e., an abstract universal, a property" (fn. omitted).

'quasi-parthood') accounts for the sense in which 3_{Ord} is *built up from* 2_{Ord}; inheriting its dependence characteristics directly from its skeleton, 3_{Ord} immediately depends on 2_{Ord}. Likewise, 3_{Ord} depends on 1_{Ord}, whose skeleton is the trivial graph,

but it does so only mediately: 3_{Ord} is built up from 1_{Ord} only in virtue of being built up from 2_{Ord}, which is in turn built up from 1_{Ord}. In the case of a limit ordinal, the picture is different. The skeleton of ω has the form

and contains the skeleton of every smaller ordinal as a proper quasi-part. But it is not *directly* produced from any of them by expansion. Thus ω should depend *mediately* on every ordinal smaller than it, but immediately on none of them—and likewise for other limit ordinals. This suggests the following dependence profile, which tracks quasi-parthood among skeletons:

(D4) $$\forall\alpha\forall\beta(\alpha \lhd_I \beta \leftrightarrow \beta = \alpha + 1) \wedge \forall\alpha\forall\beta(\alpha \lhd \beta \leftrightarrow \alpha < \beta).$$

We should be wary of assimilating quasi-parthood too closely to parthood *stricto sensu*: if the skeleton of 1_{Ord} is the sole proper part of the skeleton of 2_{Ord}, for instance, then even Weak Supplementation (the principle that if x has a proper part y, then it has also has a part z disjoint from y) is violated.[7] Nonetheless, there is a sense in which we can recapture classical mereological structure. Considering dependence *simpliciter* alone, (D4), in conjunction with the ordering requirements on ordinals, merely ensures that $\lhd$ is a well-ordering, and every (set-sized) well-ordering is order-embeddable in a model of classical extensional mereology.[8] We can justify appeal to this fact, at least in

[7] See Simons (1987: 28), Hovda (2009: 63), Varzi (2016: §3.3).

[8] *Proof*: If $\mathbb{B}$ is a Boolean algebra, then $\mathbb{B}-\{\bot\}$ is isomorphic to a model of CEM. But every well-ordered set S is order-embeddable into a $\mathbb{B}-\{\bot\}$ for some $\mathbb{B}$, as can be seen by taking $\mathbb{B}$ to be the power set algebra

part, by conceiving of (for example) the skeleton of 2_{Ord} as being the 'quasi-fusion' of the skeleton of 1_{Ord} and something else—an 'excess' component, corresponding to the graph's terminal node, which is not itself an ordinal skeleton.[9]

(Again, my official assumption here is that talk of skeletons and relations of 'extending' and 'quasi-parthood' on skeletons, like talk of abstract structures, is ultimately to be paraphrased away in higher-order terms. If it is to be taken literally, however, then the skeleton of α will, in effect, play the same role as the objectual abstract structure $\$(\mathscr{S})$ for $\mathscr{S}$ corresponding to α, and the most plausible version of the position will be one that validates (D3), for instead of objectual dependence among ordinals there will be only objectual dependence of ordinals on skeletons.)

But is (D4) the only dependence profile compatible with the bottom-up conception? It might be objected that (D4) is too strong, inasmuch as the first conjunct imposes a homogeneous treatment of limit and successor ordinals.[10] If we restrict the first conjunct to successor ordinals, we have

(D5) $\quad \forall\alpha\forall\beta(\neg\text{Limit}(\beta) \rightarrow (\alpha \lhd_I \beta \leftrightarrow \beta = \alpha + 1)) \wedge \forall\alpha\forall\beta(\alpha \lhd \beta \leftrightarrow \alpha < \beta).$

It would, however, be very *ad hoc* if some limit ordinals depended immediately on smaller ordinals and others did not, or if a limit ordinal depended immediately on some but not all ordinals smaller than it.[11] So any picture which licenses (D5) but not

over S and the embedding to be $x \mapsto \{y : y \subseteq x\}$. I am indebted to Andrew Bacon for alerting me to this connexion.

[9]Some care is required here, for it is tempting to say that the skeleton of 2_{Ord} is composed of the skeleton of 1_{Ord} 'twice over', which of course is nonsensical in a standard mereological context. See Bennett (2013) and Cotnoir (2015) for discussions of frameworks in which an object can have a part twice over. Fine's notion of 'Cantorian abstraction' (Fine 1998) may provide additional motivation for such a nonstandard picture, but we cannot explore this further here.

[10]I am indebted to an anonymous reviewer for this objection.

[11]This can be doubted: one might think that (for example) ω should immediately depend on every ordinal smaller than it, but $\omega + \omega$ should depend immediately only on the ordinals α such that $\omega \leq \alpha < \omega + \omega$, its mediate dependence on ordinals below ω being secured by way of its immediate dependence on ω. But there is a serious problem with this approach: the picture can be extended to $\omega \cdot n$ for finite n, but not to ω^2, for there is no limit $\alpha < \omega^2$ such that every ordinal strictly between $\alpha$and ω^2 is a successor. If the proponent of the view takes ω^2 (in contrast to $\omega + \omega$) to be sufficiently akin to ω to allow it to depend on every ordinal smaller than it, then similar problems will arise at ω^ω, and so on. Characterizing the class $\mathscr{F}$ of functions $f : \omega \rightarrow \text{On}$ such that $\forall\alpha(\text{Limit}\ \beta\ \&\ (\alpha \lhd_I \beta \leftrightarrow \beta < \alpha))$ if and only if $\beta = f(\omega)$ for some $f \in \mathscr{F}$ in a principled way seems a hopeless task. I thank Catharine Diehl for suggesting an option akin to this.

(D4) presumably also licenses

$$\text{(D6)}\quad \forall\alpha\forall\beta(\neg\text{Limit}(\beta) \rightarrow (\alpha \lhd_I \beta \leftrightarrow \beta = \alpha + 1)) \wedge \forall\alpha\forall\beta(\text{Limit}(\beta) \rightarrow (\alpha \lhd_I \beta \leftrightarrow \alpha < \beta)) \wedge \forall\alpha\forall\beta(\alpha \lhd \beta \leftrightarrow \alpha < \beta).$$

The differences among these principles are important: (D2)–(D6) are all incompatible with (D1), and hence with the conjunction of SBD and $\mathsf{On} = \mathsf{On}_{VN}$. (D4) and (D3) additionally rule out the quasi-Zermelo reduction $\mathsf{On} = \mathsf{On}_{QZ}$, but (D6) does not (and thus neither does the weaker, but less reflectively stable, (D5)).

But we should be wary of the claim that (D6) adequately captures the bottom-up picture. It requires a modification of the basic idea of the skeletal approach (since the skeleton of ω is certainly not the immediate extension of any other skeleton) that makes the notion of quasi-parthood still less analogous to parthood *stricto sensu*. Formal systems of mereology, axiomatized using a single primitive (either parthood or proper parthood) are not sufficiently fine-grained to distinguish between immediate parthood and parthood *simpliciter*, for reasons structurally akin to those which keep us from defining immediate dependence on the basis of dependence *simpliciter*. Nonetheless, we do have some grasp of the notions, and it seems deeply foreign to our informal mereology that an object y should have x as a proper part in both mediate and immediate ways (in the sense that there are distinct parthood chains π_1 and π_2 such that (1) there is some z such that x is a proper part of y via π_1 at least partly in virtue of x's being a proper part of z and z's being a proper part of y, and (2) there is no z such that x is a proper part of y via π_2 at least partly in virtue of x's being a proper part of z and z's being a proper part of y). Intuitions similar to those that render Strong Supplementation (the principle that, if x is not a part of y, then there is a z that is a part of y that does not overlap x) plausibly support this constraint.[12]

The proponent of (D6) might respond that there is an intuitive component involved in the bottom-up picture more important than the characterization of immedi-

[12] On Strong Supplementation, see Simons (1987: 29), Hovda (2009: 68), Varzi (2016: §3.3).

ate quasi-parthood as the converse of 'extending': the idea that the collection-at-limits principle should characterize the ordinals. To get all the ordinals from the bottom up, the defender of (D6) might claim, one must obtain limits somehow; but how is a limit to be reached save by collecting everything that has come below it? And how should collecting be understood except as taking the set-theoretical union of the things collected—a procedure which, by SBD, yields the second conjunct of (D6)?

The association between the union operation and limit ordinals is reinforced by the standard iterative picture of the set-theoretical hierarchy. Ordinarily, when we give an account of the set-theoretical universe, rather than listing the axioms of ZFC^2 or ZFCU^2, we simply give a specification through transfinite recursion on the ordinals. For the pure sets, we set $V_0 = \emptyset$; $V_{\alpha+1} = \mathscr{P}(V_\alpha)$; and, for limit δ, $V_\delta = \bigcup_{\alpha<\delta} V_\alpha$; $V = \bigcup_{\alpha\in\mathsf{On}} V_\alpha$. Similarly, given a set I of *Urelemente*, we define a hierarchy over I (and a model of *impure* set theory) as follows: $U^I_0 = I \cup \{\emptyset\}$; $U^I_{\alpha+1} = \mathscr{P}(U^I_\alpha) \cup I$; for limit δ, $U^I_\delta = \bigcup_{\alpha<\delta} U^I_\alpha$; $U^I = \bigcup_{\alpha\in\mathsf{On}} U^I_\alpha$.[13] The set-theoretical hierarchy is a paradigmatic example of an object that, by its very nature, is characterized by a function $\alpha \mapsto V_\alpha$ or $\alpha \mapsto U^I_\alpha$ obeying the collection-at-limits principle.

But it is far from clear why we should think of a limit ordinal δ *itself*, rather than the stage V_δ or U^I_δ to which it correlates, as formed by collection. Certainly it is not generally the case that every function on the ordinals, or an initial series thereof, by which a transfinite series of objects is characterized obeys the collection-at-limits principle. Consider the case of a world where space is occupied only by a point particle of one kind at location ℓ and infinitely many point particles of some different kind extending in a straight line from ℓ: denumerably many particles $a_1, a_2, \ldots$ occupy the interval parameterized by $[0.5, 1]$ (in some specified unit) on the line, the first at 0.5, the second at 0.75, and so on; these are followed by denumerably many more particles $a_{\omega+1}, a_{\omega+2}, \ldots$ occupying the interval $[1, 2)$, the first at 1, the next at 1.5, and so on. All in all, the particles form a structure of order type $\omega + \omega$. The particles spread out over $[0.5, 2)$, we can assume, are intrinsic duplicates, so we cannot appeal to particular features of

[13] Here we do not allow I to be a larger-than-set-sized collection; we shall return to this issue in §4.3.

the particles, rather than their spatial distribution, to explain their order structure. It is clear that a_ω is in no sense 'generated' by taking the union of $\{a_i : i < \omega\}$; there is no function obeying collection-at-limits that would figure in an account of the nature of the particles. To defend (D6) over (D4) is to presume that the behaviour of identity on the ordinals is more akin to that of the $\alpha \mapsto V_\alpha$ or $\alpha \mapsto U^I_\alpha$ governing the stages of the iterative hierarchy than that of the function $f : \{\alpha \in \mathsf{On} : \alpha < \omega + \omega\} \to [0.5, 2)$ characterizing the locations of the particles in our example. If the ordinals are indeed sets, then this is eminently plausible; if not, however, there is no reason to expect it to be the case.

If, then, we accept (D4), we must rule out both the $\mathsf{On} = \mathsf{On}_{VN}$ and $\mathsf{On} = \mathsf{On}_{QZ}$ reductions. In and of itself, this is a problem for reductionism, since, as was noted above, the von Neumann reduction has significant advantages in elegance and simplicity. The quasi-Zermelo reduction, although inferior, still seems preferable to more *ad hoc* options. If reductionism commits us to a candidate worse than either of these, this is a serious overall theoretical deficit.

We can, however, strengthen the argument to rule out all forms of set-theoretic reductionism about ordinals by adding some additional, independently reasonable premisses. We present the argument and then discuss the premisses:

(ODep) Either (D3) or (D4) is true.

(OHom) If any ordinal is a set, then all ordinals are sets.

(OPure) If an ordinal α is a set, then it is a pure set.

(ORank) If an ordinal α is a pure set, then $\mathsf{rank}(\alpha) \leq \alpha$.

(ODir) There are no ordinals α and β and non-ordinal pure set x such that $\alpha \lhd x$ and $x \lhd \beta$.

Proposition 4.1.1. (ODep), (OHom), (OPure), (ORank), and (ODir) in conjunction with SBD, entail that no ordinal is a set.

Proof. Assume that some ordinal is a set. By (OPure) and (OHom), every ordinal is a pure set. So ω is a pure set. By (ORank), $\mathsf{rank}(\mathrm{o}_{Ord}) = \mathrm{o}_{Ord}$, so $\mathrm{o}_{Ord} = \emptyset$. By (ODep), there is no ordinal α on which ω immediately depends, so, by SBD, ω has no ordinal as a member. But $\omega \neq \mathrm{o}_{Ord}$, so $\omega \neq \emptyset$, so there exists at least one pure set b such that $b \in \omega$. By WBD, ω depends on b. But every pure set other than $\emptyset$ itself depends, mediately or immediately, on $\emptyset$. (This result follows from WBD directly by transfinite induction.) So $\mathrm{o}_{Ord} \lhd b$ and $b \lhd \omega$, even though b is not an ordinal, which contradicts (ODir). ⊣

(ODep) has already been defended.

(OHom), the *ordinal homogeneity thesis*, is intuitively extremely plausible. If only some ordinals are sets, then which ones? The most plausible arrangement would be for there to exist some ordinal α such that all ordinals $< \alpha$ on the standard ordering are *sui generis* mathematicalia and all ordinals $\geq \alpha$ are sets. But any choice of α here is profoundly *ad hoc*.

The defence of (OPure), the *ordinal purity thesis*, divides into two cases, depending on whether contingentism or necessitism is true. I shall deal with contingentism first; it will be a standing assumption until noted otherwise that there are some things which do not necessarily exist, i.e. are not necessarily something. For simplicity, I avail myself of quantification over worlds; if needed, however, this could be paraphrased away. Given contingentism, we start from:

(4.1.3) If x is a set and x exists at w, then for all y, if $y \in \mathrm{trcl}(x)$, y exists at w.

This is usually considered a platitude about sets (Fine 1981: 180; Parsons 1983b).[14] The next premiss is:

(4.1.4) If x is an ordinal, then for all w, x exists at w.

Likewise, the claim that mathematical objects exist necessarily has very strong intuitive appeal. It is surely part of the explanation for the reaction of impatience that many

[14] We can remain neutral on whether $x \in y$ can hold at w even if x or y does not exist at w, as Bricker (1989: 387) argues; see Williamson (2013: 250 fn. 89) for discussion.

people feel when first confronted with Frege's Julius Caesar example. *Of course the Conqueror of Gaul cannot be a number*, the protest goes: he could have failed to exist (assuming contingentism), but had he (or all Gaul and all the physical universe, for that matter) not existed, it would not have made the least difference to the existence of any number.

From (4.1.3) and (4.1.4) it follows that, if any ordinal is a set, it is a *hereditarily necessary* set—a set containing only necessarily existing objects in its transitive closure. It is, of course, also a platitude about sets that every pure set is hereditarily necessary. Are there any impure sets that are hereditarily necessary? Such a set would have to have the property that every *Urelement* in its transitive closure exists neccessarily. None of the obvious necessarily existing *Urelemente* are plausible candidates for this role. *Sui generis* mathematicalia can be ruled out: the thesis that ordinals are not *themselves sui generis* mathematicalia, but sets of *sui generis* mathematicalia, merely imposes an additional, explanatorily useless layer on the antireductionist picture. *Ante rem* universals are likewise implausible; if such things exist, they must be fewer in number than Concepts, by Cantor's theorem; on the most plausible accounts, they correspond to sparse, 'natural' properties (Lewis 1983). But no collection of such properties seems a likely candidate for being an ordinal, however useful *ante rem* universals may be for other purposes.

If, on the other hand, necessitism is true, then the straightforward modal argument will not suffice: everything exists at all worlds, so (4.1.3) places no restrictions on what can figure in the transitive closure of a set that is an ordinal. But a similar argument can be made using *non-concreteness*: according to necessitism, although ordinary spatiotemporal objects exist necessarily, they are only contingently concrete. This does not mean a spatiotemporal object is abstract at a world where it is not concrete: "[i]t is a fallacy to treat 'abstract' and 'concrete' as contradictories, although they may be contraries" (Williamson 2013: 7). Yet it is intuitively as strange to claim that 5_{Ord} might not have be abstract as to claim that it might not have existed; thus, if we require that everything in the transitive closure of a set that is an ordinal be necessarily abstract,

we have reduced the candidates other than $\emptyset$ to those that were examined and found lacking earlier: *sui generis* mathematicalia and universals. Thus, whether necessitism or contingentism is true, (OPure) is on firm ground.

(ORank) reflects the idea, mentioned above, that some tight connection exists between an ordinal α and V_α (or U^I_α)—whether or not ordinals actually are sets (and whether or not a limit δ obeys collection-at-limits as V_δ and U^I_δ do). But if ordinals are sets, then if the αth rank of the hierarchy could exist 'before' (in whatever non-temporal sense of priority that is appropriate when talking about the structure of the hierarchy) the set that is the ordinal α itself existed, then the constraint that ordinals can serve as natural indices of the hierarchy would be violated.

Because any reasonable pure set reductionist account of the ordinals will provide a canonical mapping $\eta : \mathsf{On} \to \mathsf{On}_{VN}$ of the true ordinals (whatever they are) into the von Neumann ordinals, given by $\eta(\mathrm{o}) = \emptyset$ and $\eta(\alpha) = \bigcup_{\beta<\alpha} \eta(\beta)$, we can take $\mathsf{rank} : V \to \mathsf{On}$ in (ORank) to be defined simply by composing the standard $\mathsf{rank}_{VN} : V \to \mathsf{On}_{VN}$ function with the inverse of η. This is not, however, required for our purposes: all that is needed is that the rank function, whatever it is, must assign o_{Ord} to the null set and no other pure set, in accord with the iterative picture. (For the moment the iterative picture is simply presumed. In §4.3 below I shall offer a defence of the claim that, when correctly developed, it is the proper way of thinking about the set-theoretical universe.)

(ODir) corresponds to a prohibition of *ad hoc* dependence chains: the dependence of ordinals on ordinals should only happen by way of other ordinals (or, perhaps, by way of esoteric *Urelemente* such as objectual abstract structures), not by way of pure sets that are not themselves ordinals. Along with the basic datum, these requirements suffice to rule out any form of set-theoretic reductionism about ordinals.

The argument from (ODep), (OHom), (OPure), (ORank), and (ODir), and SBD to antireductionism has a direct analogue for cardinals. Here the analogues of (D3) and

(D4) are

(D3′) $$\forall\kappa\forall\lambda(\kappa \not\lhd \lambda)$$

and

(D4′) $$\forall\kappa\forall\lambda(\kappa \lhd_I \lambda \leftrightarrow \lambda = \kappa^+) \wedge \forall\kappa\forall\lambda(\kappa \lhd \lambda \leftrightarrow \kappa < \lambda).$$

(D3′)—the *pure cardinality-type conception*—is probably significantly more natural in this case than (D4′): in the absence of a distinguished order relation on a cardinal, it is difficult to see how to make sense of the notion of quasi-parthood or what could play the role of a skeleton. In the finite case, the notion that there is a unique way to build up κ^+ from κ, without imposing additional structure, makes some sense, but this breaks down at $\aleph_0$. The pure cardinality-type conception, in contrast, draws on the idea that a cardinal κ depends on an object x just in case x figures in every κ-membered class; in the next chapter, I shall discuss whether it can be derived from other principles, but it is appealing in its own right.

In any case, even with the premiss about cardinal dependence in a disjunctive form, we can argue:

(CDep) Either (D3′) or (D4′) is true.

(CHom) If any cardinal is a set, then all cardinals are sets.

(CPure) If a cardinal κ is a set, then it is a pure set.

(CSize) If a cardinal κ is a pure set, then $\mathsf{card}(\kappa) \leq \kappa$.

(CDir) There are no cardinals κ and λ and non-cardinal pure set x such that $\kappa \lhd x$ and $x \lhd \lambda$.

The arguments for (CHom), (CPure), and (CDir) parallel their analogues in the ordinal case. (CSize) is motivated by the intuition that, if cardinals have internal structure that

can be counted, that structure should not outstrip *what* they are counting—that this would be an inversion of the proper order akin to having an ordinal α occur at some rank above α.

It might be objected, however, that (CSize) objectionably rules out the cardinals generated by Scott's trick (Scott 1955)—in which the cardinal κ_{Scott} is defined as the set of all κ-membered sets of least rank. Scott's trick is certainly elegant, and $\mathsf{Card}_{Scott} = \{\{x \in V : \mathsf{card}(x) = \kappa \wedge (\forall y \in V)\ (\mathsf{card}(x) = \kappa \rightarrow \mathsf{rank}(y) \leq \mathsf{rank}(x)\} : \kappa \in \mathsf{Card}\}$ has some claim to being the best available consistent approximation to the Fregean definition of the cardinals in §68 of *Grundlagen* (Frege 1884).

In fact, however, even weakening (CSize) to allow for Scott's trick cardinals does not keep us from carrying out the antireductionist argument. Here, as in the case of the ordinals, (CSize) serves only to generate the intermediate conclusion that if o_{Card} is a pure set, then it is $\emptyset$. This is not the case if $\mathsf{Card} = \mathsf{Card}_{Scott}$, for $\mathrm{o}_{Scott} = \{\emptyset\}$. Nonetheless, from the disjunctive premiss

(CZero) If o_{Card} is a pure set, then $\mathrm{o}_{Card} = \emptyset$ or $\mathrm{o}_{Card} = \{\emptyset\}$,

we can derive an antireductionist conclusion.

Proposition 4.1.4. (CDep), (CHom), (CPure), (CZero), and (CDir) in conjunction with SBD, entail that no cardinal is a set.

Proof. Assume that some cardinal is a set. By (CHom) and (CPure), every cardinal is a pure set. So $\aleph_0$ is a pure set. By (CDep), there is no κ on which $\aleph_0$ immediately depends, so, by SBD, $\aleph_0$ has no member that is a cardinal. Now, by (CZero), either $\mathrm{o}_{Card} = \emptyset$ or $\mathrm{o}_{Card} = \{\emptyset\}$. It follows from WBD by transfinite induction that (A1) every pure set other than $\emptyset$ depends, mediately or immediately, on $\emptyset$, and (A2) every pure set other than $\emptyset$ or $\{\emptyset\}$ depends, mediately or immediately, on $\{\emptyset\}$. First consider the case in which $\mathrm{o}_{Card} = \emptyset$. Since $\aleph_0 \neq \mathrm{o}_{Card}$, $\aleph_0 \neq \emptyset$ and thus $\emptyset \lhd \aleph_0$ by (A1). Since $\aleph_0 \neq \emptyset$, there is at least one non-cardinal pure set b such that $b \in \aleph_0$, and $b \lhd \aleph_0$ by WBD. So $\mathrm{o}_{Card} \lhd b$ and $b \lhd \aleph_0$, which contradicts (CDir). Now consider the case in which $\mathrm{o}_{Card} = \{\emptyset\}$. If

$\aleph_0 \neq \emptyset$, then $\{\emptyset\} \lhd \aleph_0$ by (A2), so once again there exists a non-cardinal pure set $b \in \aleph_0$ such that $o_{Card} \lhd b$ and $b \lhd \aleph_0$, which contradicts (CDir). So $\aleph_0 = \emptyset$. But, by (CHom) and (CPure), $\aleph_\omega$ is a pure set, and by (CDep) there is no κ on which $\aleph_\omega$ immediately depends, so, by by SBD, $\aleph_\omega$ has no member that is a cardinal. In particular, $o_{Card} \notin \aleph_\omega$ and $\aleph_0 \notin \aleph_\omega$. But since $\aleph_\omega \neq \emptyset$, there exists at least one non-cardinal pure set b such that $b \in \aleph_\omega$, and $b \lhd \aleph_\omega$ by WBD. So $\aleph_0 \lhd b$ and $b \lhd \aleph_\omega$ (and, also, $o_{Card} \lhd b$ and $b \lhd \aleph_\omega$), which contradicts (CDir). ⊣

I have presented the arguments in this section using the background assumptions defended in the Introduction and Chapter I: I presume that universism about set theory is true, the theories of cardinals and ordinals are uniquely interpreted, and the reductionist positions against which the argument is directed are of the form $Ord \rightharpoonup Set$ and $Card \rightharpoonup Set$. But, as noted earlier, nothing in these arguments relies on these assumptions.

Consider, first, the case in which potentialism is true and $\hat{T}_{Card}$, $\hat{T}_{Ord}$, $\hat{T}_{Set}$, and $\hat{T}_{PureSet}$ are multiply interpreted. As noted in §1.2, it is reasonable to assume that these theories operate in tandem: any expansion of the set-theoretical universe expands the universes of ordinals and cardinals, and vice versa. In this framework, the reductionist hypotheses would take the form $Ord \overrightarrow{\rightharpoonup} PureSet$ and $Card \overrightarrow{\rightharpoonup} PureSet$. But, if this is the case, then we should expect all of (ODep), (OHom), (OPure), (ORank), (ODir), (CDep), (CHom), (CPure), (CZero), (CDir), and SBD to hold at every stage—*i.e.*, in every matched quartet of S-interpretations, in the sense of §1.2. As for multiple-interpretation reductions, the most natural choice for $Ord \overset{\text{mult}}{\rightharpoonup} PureSet$ and $Card \overset{\text{mult}}{\rightharpoonup} PureSet$ reductions would be ones on which every sequence of pure sets isomorphic to the von Neumann ordinals (or initial von Neumann ordinals, in the case of the cardials) is an S-interpretation. But the details do not matter for the argument: any $Ord \overset{\text{mult}}{\rightharpoonup} PureSet$ or $Card \overset{\text{mult}}{\rightharpoonup} PureSet$ reduction will still make the relevant premisses true in every $\mathfrak{I} \in \mathscr{I}_{Ord}$ (or $\mathfrak{I} \in \mathscr{I}_{Card}$), leading to the same result of supertruth on the natural semantics.

2. Ordinals and Cardinals by Abstraction

As I noted in the Introduction, I make no claim that the arguments of the preceding section are absolutely compelling. The background assumptions about dependence are contestable, and their plausibility has to be evaluated as part of a total theoretical package. Nonetheless, they clearly give us reason to investigate alternatives to reductionist accounts of the cardinals and ordinals: if ones can be found that are well-motivated and capable of being integrated within set theory as well as the the most appealing reductionist views, then this will make the case for antireductionism very strong.

In the case of cardinals, of course, the obvious competitor is an abstractionist account based on HP. In §3.1, I sketched how an account of mathematics following the general methodological principles defended in this work would view HP: its explanatory power and elegance give us reason to think that it characterizes the cardinals in a lawlike fashion: *what it is to be a cardinal* is to be the value of the cardinal-abstraction functor denoted by #, governed by HP. Unsurprisingly, this picture comports naturally with antireductionism: if this is the fundamental nature of the cardinals, then we should not expect them also to have additional secondary mathematical natures given by their position in the set-theoretic universe.

Can anything of this type be done for the ordinals? HP has a natural dyadic generalization, with isomorphism the analogue of equinumerosity: the Isomorphism-Type Abstraction Principle:

(ITA) $$\forall X^2 \forall Y^2 (\Omega X^n = \Omega Y^n \leftrightarrow X^2 \cong Y^2)),$$

where Ω is a type-$((e, e \to t) \to e)$ abstraction operator. Officially, we take $\cong$ as a primitive term denoting the higher-order isomorphism Relation, for reasons that will be explained in the next chapter, but for the purposes of this chapter nothing is lost by substituting a definition (in pure second-order logic) of the condition of two binary Relations' being isomorphic. ITA is the simplest abstraction principle corresponding to the *pure order-type conception of ordinals* set out above; it promises to generate an

isomorphism type abstract for every Relation whatsoever—and thus an object to serve as the order type for every well-ordered Relation.

Unfortunately, ITA is unsatisfiable on $\mathscr{L}^{2\S}$-models.

Theorem 4.2.1. ITA is unsatisfiable. (Hodes 1984: 138 fn. 16)

Proof. In order for ITA to be satisfiable on $\mathfrak{M}$, there must be an injection from equivalence classes of binary relations (*i.e.*, entities in $D_{(e,e\to t)}(\mathfrak{M})$) under isomorphism to objects (in $D_e(\mathfrak{M})$). But there are 2^κ equivalence classes of binary relations under isomorphism on a domain of first-order size κ, and by Cantor's theorem $2^\kappa > \kappa$, so there is no such injection. ⊣

Even the restriction of ITA to well-orderings, the unrestricted naïve Ordinal Abstraction Principle,

$$\text{(OAP)} \qquad \forall X^2 \forall Y^2 (\Omega X^2 = \Omega Y^2 \leftrightarrow X^2 \cong Y^2 \vee (\neg \mathrm{WO}(X^2) \wedge \neg \mathrm{WO}(Y^2))),$$

(where WO abbreviates the second-order formula expressing the condition of being a well-ordering) is unsatisfiable.

Theorem 4.2.2. OAP is unsatisfiable. (Hazen 1985: 253)

Proof. Analogous to the proof of Theorem 2, but using the fact that there are κ^+ equivalence classes of well-orderings under isomorphism on a domain of size κ. ⊣

Theorem 4.2.2 (and thus Theorem 4.2.1) can be turned into an explicit proof of inconsistency by reproducing the reasoning of the Burali-Forti paradox inside second-order logic. The property of being an ordinal and the standard operations on ordinals are definable from OAP. For every ordinal α, OAP proves the existence of an ordinal $\alpha + 1$ and it proves that $\alpha < \alpha + 1$. It further proves that, for every ordinal $\alpha, \alpha < \mho$, where $\mho = \Omega(<)$, the ordinal generated by abstracting on the ordering of the ordinals. But then $\mho + 1 < \mho < \mho + 1$.[15]

[15] Compare Hazen (1986: 286–87). I am indebted to Jeffrey Ketland for this point.

However, a way around this problem is suggested by an abstraction principle that George Boolos (1989) proposed to replace Frege's inconsistent Basic Law V,

(BLV) $$\forall X \forall Y (\in\!\!X = \in\!\!Y \leftrightarrow \forall x (Xx \leftrightarrow Yx)).$$

Boolos modifies BLV by restricting the principle of extension abstraction to Concepts *smaller in extension than the universe*; all universe-sized Concepts are mapped to a single 'rogue object', which we can denote $\infty_{\in}$.[16] His NEW V has the form

(NEW V) $$\forall X \forall Y (\in\!\!X = \in\!\!Y \leftrightarrow ((\mathrm{Big}(X) \wedge \mathrm{Big}(Y)) \vee \forall x (Xx \leftrightarrow Yx))$$

where 'Big' abbreviates the higher-order formula specifying that an injection exists from the universe to a given Concept. (For convenience, we also use 'big' and 'small' in the metalanguage with the obvious meaning.)

Applying the same technique to ITA gives us

(RITA) $$\forall X^2 \forall Y^2 (\Omega X^2 = \Omega Y^2 \leftrightarrow ((\mathrm{Big\ Fld}(X^2) \wedge \mathrm{Big\ Fld}(Y^2)) \vee X^2 \cong Y^2)),$$

where 'Fld' abbreviates a formula that, when applied to a type-$(e, e \to t)$ term, generates a type-$(e \to t)$ term for the Concept under which all and only the things in the field of the relevant Relation fall.

RITA, unlike the naïve abstraction principle ITA, is satisfiable. In order to show this, we first prove the following lemma:

LEMMA 4.2.3. Over a set of cardinality κ, where $\kappa > \lambda$, with κ and λ infinite, there are 2^{λ} distinct non-isomorphic binary relations with fields of cardinality λ.

Proof. Let A be such a set of cardinality κ, and let $\mathscr{I}_A \subseteq \mathscr{P}(\mathscr{P}(A \times A))$ be the set of isomorphism classes of λ-sized relations on A. Without loss of generality, choose some λ-sized $B \subsetneq A$. Let $\mathscr{I}_B \subseteq \mathscr{P}(\mathscr{P}(B \times B))$ be the set of isomorphism classes of relations on

[16] The term is due to Ketland (2003: 93 fn. 1).

B. It is a standard result that, on an infinite set of cardinality κ, there are 2^κ distinct non-isomorphic binary relations (Drake 1974: 67–69); so $|\mathscr{S}_B| = 2^\kappa$.[17] But consider the function f that maps an $x \in \mathscr{S}_A$ to a $y \in \mathscr{S}_B$ just in case every $z_x \in x$ is isomorphic to every $z_y \in y$; f is one-to-one and onto, so $|\mathscr{S}_A| = |\mathscr{S}_B|$. ⊣

Theorem 4.2.4. RITA is $\mathscr{L}^2$-tenable on κ just in case κ is infinite and, for all $\lambda < \kappa$, $2^\lambda \leq \kappa$.

Proof. We first handle the finite case separately; it is easy to see that RITA is unsatisfiable on a finite domain, since there are already n distinct non-Big nonisomorphic *well-orderings* of a set of size n, as well as at least one big relation that is not isomorphic to any non-big relation.[18] In the infinite case, we apply Lemma 4, with the result that the number of isomorphism types of relations with fields smaller than κ on a κ-sized structure is $\sup(\{2^\lambda : \lambda < \kappa\})$. But, by elementary cardinal arithmetic, $\kappa \geq \sup(\{2^\lambda : \lambda < \kappa\})$ if and only if, for all $\lambda < \kappa$, $2^\lambda \leq \kappa$. ⊣

This result is similar to one proved by Stewart Shapiro and Alan Weir (1999: 315–17) for NEW V: NEW V is $\mathscr{L}^2$-tenable on κ just in case κ is infinite and *regular* and, for all $\lambda < \kappa$, $2^\lambda \leq \kappa$.[19]

Corollary 4.2.5. If the Generalized Continuum Hypothesis holds, then RITA is $\mathscr{L}^2$-tenable on every infinite cardinal.

Corollary 4.2.6. If κ is inaccessible, then RITA is $\mathscr{L}^2$-tenable on κ.

Proof-theoretically, RITA is unsurprisingly fairly powerful. Let WRNA (weak relation number arithmetic) be the system whose only axiom is RITA; as usual, FA is Frege Arithmetic, the system whose only axiom is HP.

Theorem 4.2.7. WRNA interprets FA.

[17] I am indebted to Andrew Bacon for pointing me to this result.

[18] For interesting combinatorial details, see Davis (1953).

[19] The difference arises from the fact that, if for all γ, $2^\gamma \leq \kappa$, then κ is an upper bound on $\{2^\gamma : \gamma < \kappa\}$, but not on $\{\kappa^\gamma : \gamma < \kappa\}$ when κ is singular, since if κ is singular then there exists some $\gamma < \kappa$ such that $\gamma = \mathrm{cf}(\kappa)$, and $\kappa^{\mathrm{cf}(\kappa)} > \kappa$ by König's theorem.

Proof. Define $\#X = \Omega(\lambda xy.Xx \wedge x = y)$. We show that $\vdash_2 \Omega(\lambda xy.Xx \wedge x = y) = \Omega(\lambda xy.Yx \wedge x = y) \leftrightarrow X \approx Y$, where $\vdash_2$ expresses derivability in a standard axiomatization of (at least) second-order logic (such as the one given in Shapiro (1991: 65–69)).

$\Rightarrow$ Assume the antecedent. By RITΛ, either $(\lambda xy.Xx \wedge x = y) \cong (\lambda xy.Yx \wedge x = y)$ or $\text{Big}^2((\lambda xy.Xx \wedge x = y)) \wedge \text{Big}^2((\lambda xy.Yx \wedge x = y))$. In the first case, this entails $\text{Fld}((\lambda xy.Xx \wedge x = y)) \approx \text{Fld}((\lambda xy.Yx \wedge x = y))$. But $\text{Fld}((\lambda xy.Xx \wedge x = y)) \leftrightarrow X$ and $\text{Fld}((\lambda xy.Yx \wedge x = y)) \leftrightarrow Y$, so $X \approx Y$. In the second case, trivially $X \approx Y$.

$\Leftarrow$ Assume the antecedent. Either X and Y are both big or X and Y are equinumerous and both small. In the former case,$\text{Fld}([\lambda xy.Xx \wedge x = y])$ and $\text{Fld}((\lambda xy.Yx \wedge x = y))$ are both big, so by RITΛ, $\Omega(\lambda xy.Xx \wedge x = y) = \Omega(\lambda xy.Yx \wedge x = y)$. In the second case, it is provable that there is an f that bijects X and Y; $(\lambda xy.Xx \wedge x = y)ab$ if and only if $(\lambda xy.Xx \wedge x = y)f(a)f(b)$, so f is an isomorphism, so by RITΛ, $\Omega(\lambda xy.Xx \wedge x = y) = \Omega(\lambda xy.Yx \wedge x = y)$. ⊣

There are obvious definitions of the condition of being an ordinal and the natural ordering on the ordinals using RITΛ: we define $x \in \text{On}$ (or $\text{Ord}\ x$) as $\exists X^2(\neg\text{Big}\ X^2 \wedge \text{WO}\ X^2 \wedge \alpha = \Omega X^2)$ and $x <_{Ord} y$ as $\text{Ord}\,x \wedge \text{Ord}\,y \wedge \exists X^2 \exists Y^2 \exists Z^2(\alpha = \Omega X^2 \wedge \beta = \Omega X^2 \wedge X^2 \cong Z^2 \wedge Z^2$ is a proper initial segment of $Y^2)$.[20] This definition does not count the rogue object, which we term ∞_Ω, as an ordinal, which is intuitively the correct answer: since it is the abstract of multiple nonisomorphic relations, counting it as an ordinal would not be faithful to the order-type conception. (In contrast, there is a good case for counting the HP-abstract of the universal Concept, anti-zero ($\infty_\#$), as a cardinal, since those Concepts mapped to it form an equivalence class under equinumerosity.[21]) (We continue to use $\alpha, \beta, \ldots$, to abbreviate quantification restricted by Ord.)

By itself, RITΛ says nothing about whether its abstracts are sets or not. Thus, by Corollary 4.2.6, any model of pure ZFC^2 can be extended to a model of ZFC^2 + RITΛ. Moreover, for any $V_{\alpha_{VN}}$ with α_{VN} inaccessible, there exists an $f : \mathscr{P}(V_{\alpha_{VN}} \times V_{\alpha_{VN}}) \rightarrow$

[20] Like $\approx$, $\cong$, Big, and Fld, the condition of one relation's being a proper initial segment of another can be regimented in pure polyadic second-order logic; we omit the details of the formulas, which require patience but no particular creativity to write out.

[21] See Boolos (1997: 314), Rumfitt (2001).

$V_{\alpha_{VN}}$ such that $\langle V_{\alpha_{VN}}, f\rangle$ is a model of ZFC^2 + RITA and, for any well-ordering $Z \subseteq \mathscr{P}(V_{\alpha_{VN}} \times V_{\alpha_{VN}})$ such that $|\mathsf{fld}(Z)| < |V_{\alpha_{VN}}|$, f maps Z to the von Neumann ordinal isomorphic to it.[22] Thus RITA alone is guaranteed not to yield any results in the theory of ordinals *inconsistent* with the results produced by the von Neumann reduction. In the next section, we shall investigate the results of adding the principle that no ordinal is a set to RITA.

Before doing so, however, we shall examine an alternative to RITA, the Small Order-Type Abstraction Principle discussed by Roy Cook (2003: 67), which stands to OAP as RITA stands to ITA:

$$\begin{aligned}(\text{SOAP}) \quad \forall X^2 \forall Y^2 (\Omega X^2 = \Omega Y^2 \leftrightarrow (\neg \mathrm{WO}\, X^2 \vee \mathrm{Big\,Fld}\ X^2) \vee (\neg \mathrm{WO}\, Y^2 \vee \mathrm{Big\,Fld}\ Y^2) \\ \vee ((\neg \mathrm{Big\,Fld}\ X^2 \wedge \neg \mathrm{Big\,Fld}\ Y^2 \wedge \mathrm{WO}\, X^2 \wedge \mathrm{WO}\, Y^2) \wedge X^2 \cong Y^2)).\end{aligned}$$

But RITA has significant advantages over SOAP.

First of all, the generalization from ordinals to isomorphism types—what Russell and Whitehead termed 'relation numbers' (1927: *162)—is natural. In fact, in *Principia*, an elaborate arithmetic of relation numbers is developed, from which the arithmetic of ordinals falls out as a special case.[23] It might be objected that operations on isomorphism types in general do not, in fact, have a prominent place in contemporary mathematics; operations on *order types of non-well-ordered relations*, however, do. It is perfectly normal, for example, to characterize the order type of the integers as $\omega^* + \omega$, the order type of the rationals as η, the order type of a countable nonstandard model of arithmetic as $\omega + \eta(\omega^* + \omega)$, and so on. RITA makes all these order types objects of the same fundamental mathematical kind as the ordinals and allows for a homogeneous treatment of them; SOAP cannot do this. If SOAP, rather than RITA, characterizes

[22] *Proof*: Using a version of Scott's trick, first let $f(Z)$ be the set of all sets *of minimal rank* isomorphic to Z where $|\mathsf{fld}(Z)| < |V_{\alpha_{VN}}|$ but Z is *not* well-ordered; then choose some arbitrary set a that is neither a von Neumann ordinal nor one of the sets assigned in the previous step; finally, where $|\mathsf{fld}(Z)| = |V_{\alpha_{VN}}|$, let $f(Z) = a$.

[23] Russell believed that this relation arithmetic was unduly neglected, although Tarski's work on ordinal algebras (1956) represented a very general development of a fragment of it (Solomon 1989).

the ordinals, then ω is a SOAP-abstract, but ω^* is not. It may be an abstract linked to a different abstraction principle, or a set, but in any case general order-type addition will have to be defined by a cross-kind function—at a significant cost in elegance.

More importantly, RITA permits generalization to other adicities, whereas SOAP does not, since there is no natural ternary (or other higher polyadic) analogue of a well-ordering. We can introduce RITA^n,

(RITA^n) $\quad \forall X^n \forall Y^n (\Omega^n X^n = \Omega^n Y^n \leftrightarrow ((\text{Big Fld}^n(X^n) \wedge \text{Big Fld}^n(Y^n)) \vee X^n \cong_n Y^n))$,

for $n \geq 1$, where Fld^n and $\cong_n$ are the obvious n-ary generalizations of Fld and $\cong$. (Ω in RITA is Ω^2.) Model-theoretically, RITA^n for $n > 2$ is as well-behaved as RITA itself.

Theorem 4.2.8. If RITA is $\mathscr{L}^2$-tenable on κ, then so is RITA^n.

Proof. On any infinite set A, there is a bijection $f : A^n \to A \times A$ for any finite n. Coding up n-ary relations in the obvious way, it follows that for $\kappa > \lambda$, with κ and λ infinite, there are 2^λ distinct non-isomorphic n-ary binary relations with fields of cardinality λ. The remainder of the proof proceeds as in Theorem 5. ⊣

Since the singulary analogue of isomorphism is equinumerosity, RITA^1 is just

(RHP) $\quad \forall X \forall Y (\Omega^1 X = \Omega^1 Y \leftrightarrow (\text{Big } X \wedge \text{Big } Y) \vee X \approx Y))$,

but, as Ketland (2002: 92) has noted, this is provably equivalent to HP itself. Thus, acceptance of RITA^n for $n > 1$ not only yields a system of abstracts for relations more general than anything possible with SOAP, but it provides a cardinal abstraction principle for free. We can thus write # for Ω^1 and define $x \in \text{Card}$ (or Card x) as $\exists X\, x = \#X$ and $x <_{Card} y$ as $\exists X \exists Y (x = \#X \wedge y = \#X \wedge X \prec Y)$ (where $\prec$ expresses the existence of an injection from X to Y but not vice versa); we continue to use $\kappa, \lambda, \ldots$ to abbreviate quantification restricted by Card.

4.3 Integrating *Sui Generis* Cardinals and Ordinals with Set Theory

Since acceptance of the RITA^n family does not itself entail antireductionism, developing a reasonably comprehensive regimented system for the mathematical theories that interest us will require adding antireductionist axioms separately. We use 'ß' as a primitive predicate for sethood, and 'ε' as an elementhood predicate (reserving $\in$ for our informal metalanguage). The relevant axioms are:

(OAR) $\forall\alpha\,\neg ß\alpha,$

(CAR) $\forall\kappa\,\neg ß\kappa.$

Presumably, however, we should also accept the general antireductionist schema,

(GAR) $\forall x(\exists Y^n\, x = \Omega Y^n \rightarrow \neg ß x),$

from whose dyadic and singulary instances (OAR) and (CAR) follow.

But how are we to integrate RITA^n and (GAR) into a theory of sets? At first, it seems promising to continue the search for abstraction principles: if an AAR-based account of cardinals and ordinals works, could an AAR-based account of sets be feasible as well? This was what motivated Boolos to introduce NEW V. On his view, the system containing only NEW V, which he terms Frege–von Neumann Set Theory (FN)—is intended to implement the notion of *limitation of size*, one of the "two thoughts 'behind' set theory" (Boolos 1989: 103).

Limitation of size principles come in various strengths. Burgess has usefully distinguished among *maximalist* and *non-maximalist*, *qualitative* and *non-qualitative* versions: "The quantitative versions are committed to understanding 'many' and 'few' in terms of one-to-one correspondences, while the non-quantitative are not. The maximalist versions are committed to equating 'too many' with 'as many as there are objects altogether', the non-maximalist versions are not" (2005: 188).

We shall focus on the maximalist quantitative version, which appears in Boolos as

the "stronger version of limitation of size" (1989: 90):

(SLOS) Any objects whatsoever form a set just in case there is no bijection between the universe and those objects.[24]

This form of limitation of size is to be found first in von Neumann's seminal paper (1925), although Michael Hallett (1984: 286–98) has persuasively argued that it has a precursor in Cantor's notion of the Absolute.[25]

It has, however, been criticized as lacking motivation:

> The platonist is committed, on pain of contradiction, to saying that there is no collection of everything, and hence that the number of everything is absolutely infinite. Many from von Neumann on have claimed that this is the *only* absolutely infinite cardinality, but it is not clear on what ground. If all that interested us were paradox-barring, it might *suffice* to rule out just this one case, but why should we think that this is the only case that *has* been ruled out? (Potter 2004: 229)

Later on, I shall argue that SLOS *does* have a plausible motivation, based on the surrogacy principles discussed in §1.1. For the moment, however, I shall focus on the question whether FN adequately implements the idea behind SLOS.

Following Boolos (1989: 98), I term the value of the functor corresponding to the € abstraction operator in NEW V when applied to a given Concept the *subtension* of that Concept. On its own, NEW V clearly is not adequate to SLOS, for SLOS is a principle about sets and NEW V is a principle about subtensions: some bridging principle is needed to get from one to the other. The simplest approach is just to equate sets with subtensions and define elementhood as falling under a subtension:

$$\text{(4.3.1)} \qquad \text{ß}x \stackrel{\text{Df.}}{=} \exists X(x = \text{€}X);$$

$$\text{(4.3.2)} \qquad x \varepsilon y \stackrel{\text{Df.}}{=} \exists X(y = \text{€}X \wedge Xx).$$

[24] Boolos formulates the claim in terms of injections, but given Choice in the background logic this makes no difference.

[25] See the letter from von Neumann to Zermelo, 15 August 1923, quoted in Hallett (1984: 288).

Unfortunately, this proposal runs aground on the existence of the rogue object for NEW V—$€(\lambda x \,.\, x = x)$, corresponding to anti-zero for HP. As before, we use $\infty_€$ as a name for this object. (4.3.1–2), in conjunction with NEW V, implies that $\infty_€$ is a set and that $\forall x \, x \varepsilon \infty_€$, results flatly incompatible with SLOS. (And, for good measure, $\infty_€ \varepsilon \infty_€$, contradicting the Axiom of Foundation.) These problems can be avoided with modified definitions, which Boolos (1989: 99) uses:

$$ß x \stackrel{\text{Df.}}{=} \exists X(x = €X \wedge \text{Small } X), \tag{4.3.3}$$

$$x \varepsilon y \stackrel{\text{Df.}}{=} \exists X(y = €X \wedge ß y \wedge X x). \tag{4.3.4}$$

But now it is clearly the definition, and not NEW V, that does the work of sustaining SLOS.[26] An alternative proposal is to make extensionality, not size, the requirement for a subtension's counting as a set:

$$ß x \stackrel{\text{Df.}}{=} \exists X(x = €X \wedge \forall Y(€X = €Y \leftrightarrow X \equiv Y)). \tag{4.3.5}$$

But (4.3.3) makes it clear that extensionality *for sets* is not "an immediate consequence" (Boolos 1989: 102) of NEW V; it is an immediate consequence of NEW V plus a definition that either has the extensionality constraint written into it, as in (4.3.5), or itself carries the burden of the limitation of size principle that was supposed to be borne by NEW V, as in (4.3.4).

FN, thus, can hardly be called an intuitive representation of limitation of size and nothing else: it can be made to do set-theoretical work only by a choice of definitions that is far from neutral. Moreover, it suffers additional serious problems as a foundation for mathematics: the fact that it has a countable model is an obvious limitation on its ability to serve as a workable set theory. (Of course, ZFC has a countable model, but this is an artefact of the Löwenheim–Skolem theorem; FN has a countable

[26] Compare Paseau (2007: 43): "[I]t will not do to say that $[\infty_€]$ is not set-like because sets are the subtensions of small concepts and that the set-membership relation can be *stipulated* to hold between sets only. This is an *ad hoc*, entirely terminological, riposte that can always be resorted to, no matter what the similarities of the objects and relations under consideration."

model *even given a full interpretation of the second-order quantifier*.) Also, FN is mutually interpretable with PA^2, and thus provides no added proof-theoretic strength over FA (Burgess 2005: 189). These problems, taken together, suggest that there is a serious problem with the basic motivation of NEW V. NEW V is an attempt to do set theory by performing the least invasive surgery feasible on Basic Law V: the Fregean notion of sets as extensions of Concepts remains central. But one might well think, after surveying the progress of the foundations of mathematics since the 1890s, that the right lesson to draw is that Frege was simply on the wrong track: sets are just not things whose nature it is to be extensions of Concepts, although some Concepts have extensions that coincide with the members of a set.[27]

Are there other plausible abstractionist set theories to which we could turn? Some systems have succeeded in reaching greater strength, but at a significant cost in theoretical elegance. The most important are those developed by Cook (2003) and Burgess (2005: 184–210). Cook's system, as mentioned in the previous section, uses SOAP, and it has many similarities to what I shall ultimately propose. Cook's basic idea is straightforward: (almost) the entire iterative conception of set is coded up in a complex abstraction principle, NEWER V, which, in combination with some background axioms that themselves have the form of abstraction principles, (almost) yields full orthodox set theory with urelements. But Cook's system has several drawbacks. First, it incor-

[27] One objection to this line of argument involves the non-well-founded set theory developed by Peter Aczel (1988): even if the theory of Fregean extensions is not our best account of the *well-founded sets*, might it not be our best account of the *sets in general*, among which well-founded sets form only a special subclass, perhaps requiring supplemental principles of their own for a reasonable characterization? (See Jané and Uzquiano (2004) for technical work along these lines.) After all, Mirimanoff (1917: 42) seems originally to have conceived of Foundation not as a general truth about sets but as a means of distinguishing between "ordinary" and "extraordinary" (non-well-founded) sets. The central problem with this objection is that the arguments made for the existence of non-well-founded sets turn either on a fundamental motivation that is avowedly non-set-theoretic or on considerations of utility that require only non-well-founded *relations* in general, not non-well-founded elementhood. As to the first point, Aczel's anti-foundation-axiom (AFA) is most naturally formulated as a claim about graphs (1988: 10). But if the theory of graphs is interpreted as an algebraic, rather than an assertory, part of mathematics, the motivation for revising set theory vanishes: non-well-founded models that interpret the graph relation as something other than elementhood will function perfectly well. Likewise, non-well-founded mathematical structures are incontestably useful for modelling various phenomena in linguistics, philosophy, and elsewhere (Barwise and Moss 1996). This is an instance of applied mathematics that illustrates the surrogacy thesis. But nothing requires that some non-well-founded relation under study have an isomorphic surrogate in (a restriction of) $\in$; any relation over mathematicalia with the appropriate structure will do.

porates NEW V and thus inherits its problems: one technical consequence of this is that Cook's system (SOAP + NEW V + NEWER V) does not yield full second-order foundation except under restriction to pure sets (2003: 82). Second, it cannot generate an iterative hierarchy with a basis containing more urelements than can form a set; specifically, not all ordinals can be in the basis of the hierarchy, which is intuitively in conflict with the autonomy of the ordinal abstraction principle, as Cook notes (2003: 76). The system I shall develop avoids these problems.

Similarly, Burgess's Fregeanized Bernays set theory (FB) makes use of an explicit reflection axiom that does not have the form of an abstraction principle. As Burgess notes, "there is very little residue of Frege left" in the final product (2005: 213): the natural competitors of FB are thus set theories with other powerful reflection principles, and it is far from clear that it does better overall than, e.g., ZFC^2+ Philip Welch's global reflection principle (GRP_0) (Welch 2012, Welch and Horsten 2016). Moreover, Øystein Linnebo has shown that FB is "extremely sensitive [...] to the choice of primitive vocabulary" (2006: 349). It is a desideratum for a general theory with many interdefinable terms that it should be, as far as possible, 'coordinate-independent' with respect to choice of primitives, and FB fares badly in that regard.

But, in the same essay in which he introduces NEW V, Boolos provides a different account of the nature of sets, much more faithful to mathematical practice—his version of the iterative conception. This is, on his view, the *second* of the two conceptions animating set theory, independent of and complementary to limitation of size (Boolos 1989: 103).[28]

The iterative picture, Boolos contends, is "the only natural and (apparently) consistent conception of set we have" (1989: 90). But it is not wholly straightforward to formulate: it is very easy to state the picture in terms that are confessedly metaphorical, but very difficult to be certain that enough of the metaphor has been purged from the informal elucidations of the primitives. In the most unabashedly figurative terms,

[28] It is not clear that these two principles suffice: there are reasons to think that we also require some sort of reflection principle to justify large cardinal axioms (Friedman 1971, 2000; Kanamori 2003; Shapiro and Uzquiano 2008; Koellner 2009). But I cannot pursue this question further here.

sets are *constructed*, *one after the other*, in a sequence of ordered *stages*. Here is how Boolos elucidates the idea in an earlier essay:

> A set is any collection that is formed at some stage of the following process: Begin with individuals (if there are any). An individual is an object that is not a set; individuals do not contain members. At stage zero (we count from zero instead of one) form all possible collections of individuals. If there are no individuals, only one collection, the null set, is formed at this oth stage. [...] At stage one, form all possible collections of individuals and sets at stage zero. [...] At stage two, form all possible collections of individuals, sets formed at stage zero, and sets formed at stage one. [...] Keep going on in this way, at each stage forming all possible collections of individuals and sets formed at earlier stages. Immediately after all of stages one, two, three, ... , there is a stage, call it stage omega. (Boolos 1971: 18–19)

And thence to stage $\omega + 1$ and so on, through the hierarchy of stages, without end. It is important to note that, despite the use of ordinals to index the stages in the informal presentation, Boolos does not explicitly build well-foundedness into his characterization of the hierarchy at the beginning; it is supposed to fall out from his other requirements. And, of course, this use of the ordinals in itself gives no support to the claim that the ordinals *are* sets.

Given the methodological picture I have endorsed, there is no inherent reason to object to mixing an iterative, non-abstractionist account of sets with an abstractionist account of cardinals and ordinals. AAR is a thesis about the nature of abstraction principles *when those principles are powerful and do fundamental explanatory work*, not a prescription to shoehorn every kind of mathematical theory into an abstractionist framework: if no suitable abstraction principle is available, then there is no impediment to searching for other accounts of mathematical objects.

Unfortunately, there is a problem with simply taking over the standard iterative picture. RITA requires that there be as many ordinals as there are objects in the universe,

and antireductionism requires that every such object be an *Urelement*. Thus, assuming standard set theory with the axiom of Replacement, if RITA and ARP obtain, there will be no set of all *Urelemente*: the Urelement Set Axiom,

$$\text{(USA)} \qquad \exists x \forall y (y \in x \leftrightarrow \neg ß y),$$

will fail. But USA is generally thought to be integral to the iterative conception. Boolos's recipe begins "at stage zero [...] form all possible collections of individuals."[29] Since the collection of *all the individuals* is one of the possible collections of individuals, USA seems unavoidable if we follow Boolos's explanation of the hierarchy.

A survey of philosophically motivated discussions of the iterative hierarchy reveals that the assumption of USA, among systems that admit *Urelemente*, is almost ubiquitous. Among the classic presentations, those of Joseph Shoenfield (1967: 238, 1977: 322) and Hao Wang (1977: 311–12), as well as the formal system of Boolos (1989) (which we shall consider in more detail shortly) consider only pure sets; thus, the question does not arise for them. Zermelo's foundational 1930 essay implicitly imposes USA, considering only models where a set of all *Urelemente* exists at the first level (1930: 412–13); the same is true of the discussions of the intended models of set theory by Georg Kreisel (1965: 100), Dana Scott (1974: 209), James van Aken (1986: 993–95), and Michael Potter (2004: 41–47). Only Richard Montague (1965) and Charles Parsons (1977: 270) consider seriously the possibility of rejecting USA in the context of an iterative account of the universe of sets, although Montague dismisses the suggestion as yielding "a distinctly unnatural, unphilosophical, and unmathematical metatheory" (1965: 146).[30]

[29] Cf. Boolos (1989: 88–89).

[30] Parsons's discussion is of note, and is the closest to our own concerns: "In a set theory with individuals, some usual theorems about ranks, for example that for every [von Neumann] ordinal α there is a set R_α of all sets of rank $< \alpha$, require the assumption that there is a set of all individuals. It follows that there cannot be too many individuals; for example, ordinals cannot all be construed as individuals. The plausibility of this assumption depends on the intended application. It would be a piece of highly dubious metaphysics to assume that there is a set of *absolutely* all individuals, if for no other reason because it is not settled once and for all what *is* an individual. Pure mathematics should be independent of this question; for it the individuals can be an arbitrary set, class, or sometimes structure. However, below I shall assume that the individuals constitute a set" (Parsons 1977: 270). In addition, Harvey Friedman

Similarly, David Lewis rejects the possibility that concrete possible worlds might be too numerous to form a set as follows:

> How could the worlds possibly fail to comprise a set? We do say that according to the iterative conception of sets, some classes are 'too big to be sets', but this is loose talk. Sheer size is not what matters; rather, the obstacle to sethood is that the members of the class are not yet all present at any rank of the iterative hierarchy. But all the individuals, no matter how many there may be, get in already on the ground floor. So, after all, we have no notion what could stop any class of individuals—in particular, the class of all worlds—from comprising a set. (Lewis 1986: 104)

This is less a defence of the intuition than an eloquent restatement of it, but there have been more substantial attempts at motivation. The most convincing of these rely on an idea we have already discussed—*surrogacy principles*. In §1.1, I endorsed

(1.1.1) For every structure $\mathscr{S}$ studied in any mathematical discipline, there exists a structure $\mathscr{S}'$ within V such that $\mathscr{S}'$ is an adequate mathematical surrogate for $\mathscr{S}$,

and, on the basis of an analysis of surrogacy in terms of isomorphism,

(1.1.4) For every structure $\mathscr{S}$ studied in any mathematical discipline, there exists a structure $\mathscr{S}'$ within V such that $\mathscr{S}'$ is isomorphic to $\mathscr{S}$.

A number of philosophers have endorsed USA on the basis of something like (1.1.4). The general line of argument can be summed up as follows: if set theory can model only set-sized collections, but set theory, in order to carry out its functions, must be adequate for modelling all the *Urelemente*, then all the *Urelemente* must form a set. This is how the idea is formulated by Allen Hazen, defending not USA itself but "the weak hypothesis

(2004) discusses a framework with ¬USA, although his perspective is not iterativist. Menzel (1986, 2014) accepts USA but in an unusual framework, in which there is a set of all ordinals. I mention some of the problems with Menzel's abandonment of Replacement briefly below; conditions of space preclude a fuller consideration of his approach.

[...] that each world should contain a set (not a proper class) of 'physical' objects, and that they should be characterized by a set (not a proper class) of properties":

> This, I shall argue (the argument belonging as much to the philosophy of mathematics as to metaphysics of modality) is no restriction at all. Although it sounds analogous to an attempt to impose as an a priori principle a restriction to the effect that no world can be more than a trillion light-years across, to think of it in that way is to misunderstand the nature of the set/proper class distinction. Part of what is implied in taking set theory as foundational for mathematics is that every structure studied by mathematicians should be represented by a set, and the axioms of set theory have been chosen, historically, to ensure that this is so. [...] [I]t is one of the fundamental heuristic principles of set theory that no decent mathematical structure can be a proper class, but that all must be represented by sets. [...] The joint effect is that, whatever complexity can be found in the natural world, set theory, either by derivation from current axioms or by new postulation, will claim there is an isomorphic copy of it which is a set. (Hazen 1996: 35–36)[31]

If this argument is sound, it generalizes beyond physical objects to non-set abstracta. But Hazen's claim that "it is one of the fundamental heuristic principles of set theory that no decent mathematical structure can be a proper class" is clearly false: even leaving aside V itself, the constructible universe L and the other inner models studied by set theorists are certainly 'decent mathematical structures'. The notion that USA can be derived from an isomorphism constraint, however, is more promising; it is taken up by McGee:

> Absent a comprehensive ontological theory, what motivates the *Urelement* Set Axiom is not the hypothesis that the *Urelemente* are few, but the thesis that the pure sets are many. So rich and varied are the structures to be

[31] Cf. Hazen (2004: 328–29).

> found among the pure sets that, whatever the *Urelemente* may turn out to be, an isomorphic copy of their structure can be found within the pure sets. (McGee 1997: 53)[32]

McGee does not give an explicit motivation for this isomorphism requirement, but something like the surrogacy argument seems to be in the background.

But why should we think that (1.1.4) requires USA? This will trivially be the case if the informal notion of isomorphism between structures in (1.1.4) is made rigorous using isomorphism *as it is defined in set theory—i.e.*, as a relation on sets. But, in a framework with primitive higher-order quantification, this is an unnatural way to understand the notion: the natural approach is simply to treat a particular, adicity-specific notion of isomorphism (*e.g.*, on type-$\overline{(e, e \to t)}$ Relations) directly (*e.g.*, as a type-$\overline{((e, e \to t), (e, e \to t) \to t)}$ Superrelation). On this *logical* notion of isomorphism, (1.1.4) is satisfied just in case there is an injection from U, the universal class, to V, the class of all pure sets. In other words, what respect for the representational prowess of set theory requires is not USA but the weaker Universal Injection Axiom,

(UIA) $$(\lambda x \,.\, x = x) \prec (\lambda x \,.\, ß_{\mathsf{Pure}} x),$$

where $\prec$ expresses the higher-order condition of being an injection and $ß_{\mathsf{Pure}}$ abbreviates a formula expressing the condition of being a pure set, defined using ß and standard set-theoretic resources. And UIA, unlike USA, does not conflict with the conjunction of the RITA^n family with (OAR) and (CAR).[33]

These considerations undercut the motivation for USA, and thereby the challenge it

[32] Cf. McGee (2006: 97).

[33] McGee offers another argument for USA: the "Categoricity Theorem" that "Any two models of second-order ZFCU + the *Urelement* Set Axiom with the same universe of discourse have isomorphic pure sets. In particular, any two models of second-order ZFCU + the *Urelement* Set Axiom in which the first-order variables range over everything have isomorphic pure sets" (1997: 55). However, as Rayo and Uzquiano (1999: 324 fn. 1) note, USA is not in fact required: "A little inspection of the proof reveals that what is required for the result to be provable is that one can prove that there is a 1–1 correspondence between the universe of pure sets and the universe of discourse"—i.e., UIA. In any case, McGee's motivation for seeking the categoricity result depends on an extreme version of set-theoretic pluralism, with strongly structuralist motivations, which I reject for the reasons given in §1.2.

poses to (OAR) and (CAR), but they do not themselves help with the positive part of the *sui generis* account: they do not show how to develop an iterative account of set theory in the absence of USA. In order to do this, I present a formal system incorporating a hierarchy of levels of sets, indexed by ordinals, based on the system presented by Boolos (1989: 91).

Boolos's original system (which I shall term Boolos Iterative Set Theory, BI) is a theory of pure sets, from which second-order Zermelo set theory (Z^2) can be derived. He presents it as a schematic theory, but I shall use full second-order logic. It is a two-sorted theory: $\rho, \rho_1, \rho_2, \ldots$, range over stages, $x, y, z, \ldots$, over sets. The primitives are elementhood (ε), an ordering on stages ($\sqsubset$), and the 'occurs at' relation between an object and a stage (@); $x \blacktriangleleft \rho$ ('x occurs before stage ρ') is defined as $\exists \rho_1 (\rho_1 \sqsubset \rho \wedge x@\rho_1)$. The axioms are:

(BI1) $$\forall \rho_1 \forall \rho_2 \forall \rho_3 (\rho_1 \sqsubset \rho_2 \wedge \rho_2 \sqsubset \rho_3 \rightarrow \rho_1 \sqsubset \rho_3);$$

(BI2) $$\forall \rho_1 \forall \rho_2 \exists \rho_3 (\rho_1 \sqsubset \rho_3 \wedge \rho_2 \sqsubset \rho_3);$$

(BI3) $$\exists \rho_1 (\exists \rho_2 \, \rho_2 \sqsubset \rho_1 \wedge \forall \rho_2 (\rho_2 \sqsubset \rho_1 \rightarrow \exists \rho_3 (\rho_2 \sqsubset \rho_3 \wedge \rho_3 \sqsubset \rho_1)));$$

(BI4) $$\forall x \exists \rho \, x@\rho;$$

(BI5) $$\forall x \forall \rho (x@\rho \leftrightarrow \forall y (y \varepsilon x \rightarrow y \blacktriangleleft \rho));$$

(BI6) $$\forall X (\exists \rho \forall x (Xx \rightarrow x \blacktriangleleft \rho) \rightarrow \exists y \forall x (x \varepsilon y \leftrightarrow Xx));$$

(EXT) $$\forall x \forall y (x = y \leftrightarrow \forall z (z \varepsilon x \leftrightarrow z \varepsilon y)).$$

The extensionality assumption is given separately, since, according to Boolos, "it may be 'analytic', or 'analytic-whatever-it-may-mean-to-say-so' that different sets have different members, but that they do is not actually guaranteed by the iterative conception, properly so-called" (1989: 93).

BI1–3 ensure that the hierarchy of stages is transitive, has no maximal stage, and has a non-initial stage without an immediate predecessor. BI4 expresses the fact that the hierarchy is total: it contains every pure set. BI5 says that a set is located at a level ρ just

in case all its elements are located somewhere below ρ. The hierarchy is, as expected, cumulative: if a set occurs at stage ρ, then it also occurs at every stage beyond ρ. BI6 is a specification principle, akin to the *Aussonderungsaxiom* (which can be derived from BI). Although there is no explicit version of the Axiom of Choice, set-theoretic Choice is derivable if there are choice principles in the background logic.[34] Replacement cannot be derived from BI, but if we add a regimented version of SLOS in by hand,

$$\text{(LOS)} \qquad \forall X(\neg \text{Big } X \leftrightarrow \exists y \forall x(x \varepsilon y \leftrightarrow Xx)),$$

we can obtain it (Boolos 1989: 104).[35]

Our reasons for rejecting NEW V as a way of implementing SLOS were not, of course, objections to SLOS itself, or to Replacement. In fact, the same general picture that motivates (1.1.4) can motivate Replacement (and, indirectly, SLOS itself). Not only is it mathematically fruitful, but, as Kanamori (2012: 47) remarks, it corresponds to "a certain kind of *indifference to identification*, in set theory and in modern mathematics generally". It was this indifference that was emphasized in §1.1: not only is set theory sufficiently powerful to represent any structure, up to isomorphism, but it does not discriminate in the way it treats isomorphic structures. Hence, if the objects in the carrier class of one structure $\mathscr{S}$ form a set, then so do the objects in those of an isomorphic class $\mathscr{S}'$. But whether some objects form a set should not depend on what kind of structure over them is under consideration: this leads to the idea that, if some things form a set, then any other things equinumerous with them should also form a set. Another generalization, from the images of bijections to the images of any functions whatsoever, yields Replacement. From there, it is a short step to the idea that not only does size alone determine sethood, but it does so in the most direct way possible given UIA: SLOS. This line of thought is not irresistible, of course, but it is very natural—and it explains why systems that are forced to limit Replacement, such as the one in Menzel (2014), seem artifically constrained.

[34] See Pollard (1988, 1989) and Potter (2004: 257).
[35] A reflection principle would also suffice: see Scott (1974: 213).

Taking all these elements together, we have a system of pure set theory from which all the axioms of ZFC² can be derived. But, if we merely introduce *Urelemente* into BI in the most straightforward fashion, replacing BI5 with

(BI5′) $$\forall x \forall \rho (x@\rho \leftrightarrow \neg ß x \vee \forall y (y \varepsilon x \rightarrow y \blacktriangleleft \rho))$$

and EXT with the uncontroversial

(EXT′) $$\forall x \forall y (ß x \wedge ß y \rightarrow (x = y \leftrightarrow \forall z (z \varepsilon x \leftrightarrow z \varepsilon y))),$$

and adding the obvious principle

(SET-ELEM) $$\forall x \forall y (x \varepsilon y \rightarrow ß y),$$

we get a system that proves USA trivially.

Instead, we shall develop Modified Boolos Set Theory (MB) by *relativizing* the notion of a hierarchy to a *basis* (a less-than-universe-sized collection of *Urelemente*).[36] We work in a second-order language with ß, ε, Ω^n for each $n \geq 1$, and the type-$(e, e, (e \rightarrow t) \rightarrow t)$ operator @ (taking two first-order terms and a singulary predicaet term and generating a formula). The intended interpretation of $x@^X\alpha$ is: the object x occurs at stage α of the hierarchy over the basis X. (We let $x@^X y$ be false whenever y is not an ordinal.)

Although the relativization requirement means that $x@^X\alpha$ holds only when X has less-than-universe-sized extension, there is nonetheless a perfectly coherent conception of *all the things at level α in the hierarchy*. It is just that this is a proper class, and must thus officially be discussed in higher-order terms. Just as we can introduce U as an abbreviation for $(\lambda x \,.\, x = x)$, we can introduce U_α for each α as an abbreviation of $(\lambda x . \exists X\, (\text{Basis } X \wedge x@^X\alpha))$; since every object occurs at some level in the hierarchy over some basis (in fact, in many hierarchies over many bases), the principle $U = \bigcup_{\alpha \in \mathsf{On}} U_\alpha$

[36] The term is inspired by Zermelo's notion of the "*Basis des Gesamtbereiches*" (1930: 410).

comes out true (when given a suitable higher-order formulation), underscoring the claim of MB to be a genuinely iterative account of the set-theoretical universe.

Our first axioms are the RITA^n family and (GAR), and we use the definitions of being an ordinal and a cardinal (and the corresponding orderings) given before. We then add EXT′ and SET-ELEM. Our first distinctive axiom says that every stage is an ordinal:

(MB1) $$\forall X \forall x \forall y (x@^X y \rightarrow \exists \alpha\, y = \alpha).$$

The axiom BI1 is no longer needed, since transitivity is guaranteed by the properties of $<$. We still wish, however, to ensure that the sequence of ordinals extends beyond ω and that there is no terminal ordinal, which is not ensured by RITA itself (since RITA has models of size $\aleph_0$ and, if GCH holds, of successor size). Thus we keep BI3; BI2, however, can be weakened, since $<$ is guaranteed to be connected:

(MB2) $$\forall \alpha \exists \beta (\alpha < \beta);$$

(MB3) $$\exists \alpha (\exists \beta\, \beta < \alpha \wedge \neg \forall \beta (\beta < \alpha \rightarrow \exists \gamma (\beta < \gamma \wedge \gamma < \alpha))).$$

We define Basis X as $\forall x (Xx \rightarrow \neg \exists y (y \varepsilon x)) \wedge \neg$Big X; i.e., a Concept is a basis just in case its extension is smaller-than-universe-sized and contains only *Urelemente* and $\emptyset$. We can now relativize BI4, replacing it with the claim that everything occurs at some stage *in some subhierarchy*:

(MB4) $$\forall x \exists X \exists \alpha (\text{Basis } X \wedge x@^X \alpha).$$

We define $x \blacktriangleleft^X \alpha$ as $\exists \beta (\beta < \alpha \wedge x@^X \beta)$ and relativize BI5,

(MB5) $$\forall X (\text{Basis } X \rightarrow \forall x \forall \alpha (x@^X \alpha \rightarrow Xx \vee \forall y (y \varepsilon x \rightarrow y \blacktriangleleft^X \alpha))),$$

and BI6,

(MB6) $\quad \forall X(\text{Basis } X \to \forall Y(\exists\alpha\forall x(Yx \to x \blacktriangleleft^X \alpha) \to \exists y(\forall x(x\varepsilon y \leftrightarrow Yx) \wedge \exists\alpha\, y@^X\alpha))).$

Finally, we keep LOS in order to provide the strength needed for Replacement.

First, we prove consistency relative to the existence of a standard model of ZFCU^2_∞ (i.e. $\mathsf{ZFCU}^2 + \neg\text{USA}$):

THEOREM 4.3.1 Every $\mathscr{L}^2$-model $\mathfrak{M}_0 = \langle M, [\![ß]\!], [\![\varepsilon]\!]\rangle$ of ZFCU^2_∞ can be extended to an $\mathscr{L}^2$-model $\mathfrak{M} = \langle M, [\![ß]\!], [\![\varepsilon]\!], [\![@]\!], [\![\Omega]\!]\rangle$ of MB.

Proof. We construct $\mathfrak{M}$ as follows. $M, [\![ß]\!], [\![\varepsilon]\!]$ remain the same; define $\Xi = M \setminus [\![ß]\!] \cup \emptyset$. Let $\alpha, \beta, \ldots$, range over the von Neumann ordinals of M, denoted by $\mathsf{On}_{VN(M)}$. Choose an appropriate isomorphism-type abstraction function for $[\![\Omega]\!]$ (since $|M|$ is inaccessible, we are guaranteed that this is possible by Corollary 7); let On_M be $\{x : \exists z \subseteq \mathscr{P}(M \times M)(\|z\| < \|M\| \wedge \mathsf{wo}(z) \wedge x = [\![\Omega]\!](z)\}$. For $x \subseteq \Xi$, define $\hat{x}_0 = x$; $\hat{x}_{\alpha+1} = \mathscr{P}(\hat{x}_\alpha)$; for limit λ, $\hat{x}_\lambda = \bigcup_{\alpha<\lambda} \hat{x}_\alpha$; $\hat{x} = \bigcup_{\alpha\in\mathsf{On}_{VN(M)}} \hat{x}_\alpha$. Define $[\![@]\!] \subseteq \mathscr{P}(M) \times M \times M$ such that $\langle x, \alpha, z\rangle \in [\![@]\!]$ just in case $x \subseteq \Xi$, $\|x\| < \|M\|$, $\alpha \in \mathsf{On}_M$, and $z \in \hat{x}_\alpha$. Proceed similarly for the polyadic RITA^n cases, using an appropriate pairing function. ⊣

COROLLARY 4.3.2 MB is consistent relative to ZFCU^2_∞ + "there exists an inaccessible cardinal".

COROLLARY 4.3.3 MB is consistent relative to ZFC^2 + "there exists an inaccessible cardinal".

Proof. By a trivial variation on the standard model construction used to show the equiconsistency of ZFC and ZFCU_∞, for which see Hazen (2004: 327–28). ⊣

THEOREM 4.3.4 $\mathsf{MB} \vdash \mathsf{ZFCU}^2_\infty$.

Proof. (Sketch.) The proofs follow those used by Boolos (1989: 102–4) but with considerable simplification due to the presence of LOS. Extensionality is explicit; Boolos's

proofs of Foundation, Infinity, Power Set, and Null Set go through with trivial modifications. From Los we derive Replacement, and thence the *Aussonderungsaxiom* by standard techniques. Choice is derivable by way of the choice principles in the background logic, as noted above. This leaves Pairing and Union. We note that, when the basis is empty, MB reduces to BI. From this, it follows trivially that $\mathsf{MB} \vdash (\mathsf{ZFC}^2)^{Pure}$, i.e. ZFC^2 with quantifiers explicitly relativized to pure sets; we can then apply Pairing and Union on pure sets in conjunction with Replacement to obtain unrestricted Pairing and Union. ⊣

4.4 Ordinals, Cardinals, and Natural Numbers

Together, (OAR) and (CAR) represent our central claim—the rejection of set-theoretic ontological reductionism for cardinals and ordinals. But they tell us nothing about whether an $Card \rightharpoonup Ord$ or $Ord \rightharpoonup Card$ reduction thesis holds, nor do they tell us anything about how the natural numbers fit in the taxonomy of mathematical objects. Nor does MB as a whole answer these questions.

With respect to $Card \rightharpoonup Ord$ and $Ord \rightharpoonup Card$ reductions, the question is a special case of one discussed by Cook and Philip Ebert (2005: 125): under what conditions do the objects falling in the range of a functor $\mathfrak{f}_1$ figuring in a legitimate abstraction principle $\mathscr{A}_1$ and those falling in the range of a functor $\mathfrak{f}_2$ figuring in a legitimate abstraction principle $\mathscr{A}_2$ coincide? Some care is required, however, in posing the problem. They attempt to canvass various general answers to the question from the standpoint of neologicism, motivating intuitions by considering the finite version of HP:

(FHP) $$\forall X \forall Y (\hat{\#}X = \hat{\#}Y \leftrightarrow (\neg \mathrm{Fin}\ X \wedge \neg \mathrm{Fin}\ Y) \vee X \approx Y),$$

where Fin abbreviates a formula defining finitude. They consider it a desideratum that the non-rogue abstracts of the abstractor picked out by $\hat{\#}$ be identical to the non-rogue abstracts of the #-abstractor in HP (given the truth of the $Nat \rightharpoonup Card$ reduction). For a proponent of AAR rather than a neologicist, however, the question does not arise

in this form: She endorses HP because of its explanatory power, given her antecedent conception of the cardinals and her understanding of the ordinary language expression 'the number of'. She does not think that HP defines the # operator; instead, she takes the principle to be a substantive claim, true because of nontrivial facts about the nature of cardinal numbers. As such, it is meaningful to ask about the objects in the range of the # operator. In contrast, it is not clear how to take the $\hat{\#}$ operator in FHP. It can be given an informal interpretation as 'the cardinal number of ... (for a finite Concept)', but then the question of its values, for finite Concepts, is trivial: its range has been fixed parasitically in such a way that it overlaps with HP. If, however, nothing is said about what $\hat{\#}$ *means*, then FHP is not a fully interpreted sentence; there are infinitely many assignments of abstractors to $\hat{\#}$ that make FHP true, and no guidance has been given as to which one is intended. Because an abstraction principle *for natural numbers alone*, rather than for cardinals, is not particularly explanatorily powerful, there is no reason to think that there is any privileged joint-carving abstractor obeying FHP that is particularly tightly linked to the nature of the natural numbers, whereas there is reason to believe that there is such an abstractor for HP. Because AAR treats élite and non-élite abstraction functors differently, it has the consequence that we should not be confident in the existence of general solutons to Cook and Ebert's question: we shall have to consider each case on its merits.

So how do $Card \rightharpoonup Ord$ and $Ord \rightharpoonup Card$ reductions fare? Only $Card \rightharpoonup Ord$ is a reasonable candidate. The sole non-*ad hoc* reduction in the other direction is a $Card \rightleftharpoons Ord$ identification according to which the ordinals under their natural ordering just are the cardinals under theirs, but this results in highly implausible consequences: $\omega_{Ord} = \aleph_{0\,Card}$, as expected, but $\omega + 1_{Ord} = \aleph_{1\,Card}$. In the $Card \rightharpoonup Ord$ direction, however, the natural reduction hypothesis is that the cardinals are all and only the initial ordinals (the motivating idea of the $Card_{VN} \rightharpoonup Ord_{VN}$ reduction). But if this is the case, then every cardinal is essentially characterized by two distinct abstraction principles. Whether or not this is objectionable in cases where the abstraction principles have the same adicity, it is very hard to motivate in the case of a pair consist-

ing of a dyadic and a singulary abstraction principle. Cook and Ebert (2005: 128–29) try to motivate general solutions to the question by focussing on *overlap* (the Concepts falling in the intersection of the inverse image of the non-rogue values of the functor $\mathfrak{f}_1$ and the non-rogue values of the functor $\mathfrak{f}_2$). But there can be no such overlap when the abstraction operators are of different adicity (although one runs up against the limits imposed by typing in trying to state this fact, since $X^n \equiv Y^m$ is not false but simply ill-formed if $m \neq n$).

We thus have good grounds to reject $Card \rightharpoonup Ord$. In fact, we should presumably accept a general nonidentity schema:

(NID) $\forall X^n \forall Y^m \, \Omega X^n \neq \Omega Y^m$, when $n \neq m$.

The $Nat \rightharpoonup Card$ reduction thesis, *cardinalism*, is a more interesting question. It has great intuitive appeal; after Frege (1884), many philosophers have simply assumed that every natural number just is the corresponding finite cardinal. But there have been dissenters—notably Husserl (1891)—and the reduction has recently been attacked by Linnebo (2009).

One of Linnebo's objections is particularly important: he contends that "if the cardinal conception had been correct, then certain special numbers"—zero and the transfinite cardinals—"would have been obvious and unproblematic in a way they are not" (2009: 226). To be sure, anyone who maintains that the natural numbers are finite cardinals, and thus that our informal theory $\hat{T}_{Nat}$ is a fragment of the theory $\hat{T}_{Card}$, must hold that pre-Cantorian mathematicians were engaged in the enterprise of exploring one part of the mathematical world in nearly complete ignorance of the fact that it was only an antechamber to a much vaster and stranger realm. Perhaps he must similarly hold that many children learning arithmetic are (and mathematicians before the 1500s in Europe and the 600s in India were) ignorant even of one of the great features of the antechamber's architecture, the cardinal number zero. But this poses a problem for cardinalism only on an implausibly internalist and narrowly belief-based metaseman-

tics, at odds with the picture behind AAR: we have good reasons to think that most scientific and philosophical terms latch onto natural kinds, regardless of the details of speakers' beliefs about the characteristics of the objects those terms pick out; all that is needed is that use facts should succeed in constraining the meaning of a term sufficiently to bring it *reasonably close* to some natural kind.

Moreover, there is a strong direct argument for the $Nat \rightharpoonup Card$ reduction. We start from a datum about the natural numbers: whatever they are, they are the referents of canonically formed natural-language numeral terms. We need not develop a precise definition of what is required for a mathematical notation system to be canonical (relative to a language); the rough idea is that a notation system $\mathscr{Q}$ is canonical relative to $\mathscr{L}$ just in case a fully competent speaker of $\mathscr{L}$, after coming to know a piece of mathematical information by apprehending a true sentence $\ulcorner\phi(\dot{n})\urcorner$ of $\mathscr{L}$ (for instance, at the end of a proof), where $\dot{n}$ is a term of $\mathscr{Q}$, is not then in a position meaningfully to ask the question: $\ulcorner$But which number is $\dot{n}$?$\urcorner$ Canonical terms of this type provide what Kripke (1992) called 'buck-stopping identifications'; they also seem to play a mediating role in some of our *de re* knowledge about mathematicalia.[37]

In English both the word sequence 'zero', 'one', 'two', 'three', etc., and the digit sequence '0', '1', '2', '3', etc., are presumably canonical in this way; here I shall use the former.[38] We thus have:

(4.4.1) For every natural number n, there exists a canonical expression $\dot{n}$ of English ('zero', 'one', 'two', 'three', etc.) that denotes n.

Of course, (4.4.1) will not be acceptable to nominalists who adopt a paraphrase strategy on which numerals are something other than singular terms, but it should be unobjectionable to everyone who accepts the version of realism presupposed here.

The next premiss is similarly anodyne:

[37] On Kripke's ideas, see Steiner (2011). For related discussion, see Kaplan (1968), Shapiro (1983), Auerbach (1994), Horsten (2005).

[38] It is sometimes objected that the former words do not truly make up a recursive numeral notation, since at some point the series 'million', 'billion', 'trillion', and so on, runs out. But we can simply use 'thousand thousand thousand . . . '; there is no requirement that a number denoted by a term in a canonical system $\mathscr{Q}$ not also be denoted by some very similar but non-canonical term in the language.

(4.4.2) Where there are infinitely many Φs, the expression ⌜the number of Φs⌝ denotes a cardinal.

The most important premiss is a uniformity assumption, which will be defended in detail below:

(4.4.3) If any expression of the form ⌜the number of Φs⌝ denotes a cardinal, then every expression of the form ⌜the number of Φs⌝ denotes a cardinal.

From (4.4.1–3) we have:

(4.4.4) Every expression of the form ⌜the number of Φs⌝ denotes a cardinal.

The next premiss is similarly straightforward:

(4.4.5) For every natural number n, there is a true sentence of the form ⌜$\dot{n}$ is the number of Φs⌝, where $\dot{n}$ is a canonical expression denoting n.

We can follow the Fregean method for proving the infinity of the natural number series and use ⌜the number of natural numbers less than $\dot{n}$⌝ for Φ, but any other predicate that generates a true sentence will do just as well; the Fregean choice has the advantage of being free of semantic vocabulary and thus avoiding questions about how to limit truth-predicate-introduction in the metalanguage to prevent paradox.

The remainder follows by basic grammar and disquotational principles:

(4.4.6) In every sentence of the form ⌜$\dot{n}$ is the number of Φs⌝, where $\dot{n}$ is a canonical expression denoting n, the copula expresses literal identity.

(4.4.7) For every natural number n, there is a true sentence of the form ⌜$\dot{n}$ is the number of Φs⌝, where $\dot{n}$ is a canonical expression denoting n, the expression ⌜the number of Φs⌝ denotes a cardinal, and the copula expresses literal identity.

(4.4.8) If there is a true sentence of the form ⌜**a** is **b**⌝ in which **a** is a singular term denoting **b**, the copula expresses numerical identity, and **b** denotes something that is F, then a is F.

(4.4.9) For every natural number n, n is a cardinal.

Clearly, the anti-cardinalist's best option is to challenge (4.4.3).

The are two plausible manoeuvres. The anti-cardinalist might accept a $Nat \rightharpoonup Ord$ reduction and claim that, when there are infinitely many Φs, ⌜the number of Φs⌝ denotes a transfinite cardinal, but when there are only finitely many Φs, ⌜the number of Φs⌝ denotes a finite ordinal; or she might follow the anti-reductionist line, on which ⌜the number of Φs⌝ in the finite case denotes neither a cardinal nor an ordinal but a *sui generis* mathematicale. Each of these two positions, however, faces the same basic challenge. The anti-cardinalist, regardless of whether she adopts ordinalism or the *sui generis* position, must take one of two positions about 'the number of': either it is ambiguous, picking out different higher-order functions depending on context, or it picks out a single higher-order function whose values are gerrymandered among the natural numbers and the transfinite cardinals.

Let us consider the first case. We can use #, in keeping with our practice, to represent the function taking Concepts to cardinals; we introduce £ for the function taking Concepts with finite extension to the natural numbers, on the assumption that they are distinct from the finite cardinals. Thus the anti-cardinalist considers

(4.4.10) The number of regions of Gaul is three,

(4.4.11) The number of days in the year World War I began is three hundred and sixty-five,

(4.4.12) The number of primes is aleph-nought,

and

(4.4.13) The number of complex numbers is beth-one

to have the following logical forms:

(4.4.14) $£R = 3_{Nat}$,

(4.4.15) $£D = 365_{Nat}$,

(4.4.16) $\#P = \aleph_{0\,Card}$,

(4.4.17) $\#C = \beth_{1\,Card}$.

But this immediately leads to a difficulty in construing sentences such as

(4.4.18) The number of regions of Gaul is less than the number of days in the year World War I began,

(4.4.19) The number of days in the year World War I began is less than the number of primes,

(4.4.20) The number of primes is less than the number of complex numbers.

The anti-cardinalist's nonuniform analysis of 'number of' tells us nothing about how to understand 'less than'.

One natural response would be for the anti-cardinalist to maintain that *two* relations are being picked out here, which we can call $<_{Nat}$ and $<_{Card}$. The field of $<_{Nat}$ is a subclass of $\mathbb{N} \times \mathbb{N}$, the field of $<_{Card}$ a subclass of $\mathsf{Card} \times \mathsf{Card}$. Since $\mathbb{N}$ and Card are disjoint, nothing is an element of the fields of both $<_{Nat}$ and $<_{Card}$. This gives us a straightforward interpretation of (4.4.18) and (4.4.20) as

(4.4.21) $\pounds R <_{Nat} \pounds D$

and

(4.4.22) $\#P <_{Card} \#C$,

respectively. But for (4.4.19), we encounter failure, as both

(4.4.23) $\pounds D <_{Nat} \#P$

and

(4.4.24) $\pounds D <_{Card} \#P$

are false. It would be possible to make (4.4.19) come out true by positing a *third* interpretation for 'less than', which would obtain in contexts in which Concepts with finite extension are compared to Concepts with infinite extension, thus:

(4.4.25) $\pounds D <_{cross} \#P$.

Let $\vartheta : \mathbb{N} \to \text{Card}$ be the function that maps a natural number to the corresponding cardinal number under the standard ordering. (In a more formal presentation, if we follow Frege in requiring functions to be total, we can consider the values of ϑ for arguments not in $\mathbb{N}$ to be unimportant). Now $x <_{cross} y$ will obtain just in case $\vartheta(x) <_{Card} y$.

This approach, however, fails to explain the analytic validity of the inference from (4.4.18) and (4.4.19) to

(4.4.26) The number of regions of Gaul is less than the number of primes.

Treating the inference as an enthymeme with the transitivity of both $<_{Nat}$ and $<_{cross}$ as background premisses, we have

(4.4.27) $\forall x \forall y \forall z (x <_{Nat} y \wedge y <_{Nat} z \to x <_{Nat} z)$,

(4.4.28) $\forall x \forall y \forall z (x <_{cross} y \wedge y <_{cross} z \to x <_{cross} z)$,

(4.4.29) $\pounds R <_{Nat} \pounds D$,

(4.4.30) $\pounds D <_{cross} \#P$,

(4.4.31) $\pounds R <_{cross} \#P$,

which is invalid.

In view of these problems, it seems that the anti-cardinalist is better off giving 'less than' a uniform, *amalgamated*, reading, as follows:

(4.4.32) $x <_{amalg} y \leftrightarrow (\text{Nat } x \wedge \text{Nat } y \wedge x <_{Nat} y) \vee (\text{Card } x \wedge \text{Card } y \wedge x <_{Card} y)$
$\vee (\text{Nat } x \wedge \text{Card } y \wedge \vartheta(x) <_{Card} y) \vee (\text{Card } x \wedge \text{Nat } y \wedge x <_{Card} \vartheta(y))$,

where Nat is a natural number primitive.

This need not be seen as a definition, but it is presumably intended to track metaphysical priority: $<_{Nat}$ and $<_{Card}$ are basic, of a piece with $\mathbb{N}$ and Card themselves, whereas $<_{amalg}$ is a less fundamental bridging relation. The use of $<_{amalg}$ recovers the expected results for the truth of (4.4.19) and the validity of the inference from (4.4.18–19) to (4.4.26), which now takes the form:

(4.4.33) $\pounds R <_{amalg} \pounds D$

(4.4.34) $\pounds D <_{amalg} \#P$,

(4.4.35) $\pounds R <_{amalg} \#P$,

But $<_{amalg}$ is a poor choice for 'less than', as it fails to totally order its own range: we have $2_{Nat} \neq 2_{Card}$, but neither $2_{Nat} <_{amalg} 2_{Card}$ nor $2_{Card} <_{amalg} 2_{Nat}$. (Presumably 'greater than' is simply the converse of 'less than', whatever 'less than' means, so we also have $2_{Nat} >_{amalg} 2_{Card}$.)

Of course, few important relations in mathematics totally order everything. None of $x <_{Card} y$, $y <_{Card} x$, and $x = y$ will obtain when x and y are distinct non-cardinals. Yet we normally say that $<_{Card}$ obeys trichotomy, either treating it as undefined over non-cardinals or (if we follow the Fregean line and insist on totality) tacitly excluding irrelevant objects by a context-dependent restriction of the quantifier. So it might be assumed that, through judicious reliance on context and sortal restrictions, the proponent of the amalgamation strategy can avoid ever having to deal with an ordinary language failure of trichotomy for 'less than'.

But the following sentence seems like a platitude:

(4.4.36) If anything whatsoever is less than the number of real numbers, then if it is neither less than nor greater than the number of seeds on the plate, it is the number of seeds on the plate.

No contextual restriction will help here unless it is a restriction on 'anything whatsoever'—whose explicit purpose is to lift quantifier restrictions.

Confronted with these problems, the anti-cardinalist might posit that 'the number of' always picks out the same higher-order function, but this function is gerrymandered across mathematical kinds. Let us use $\mathfrak{b}$ to represent this function: where X has infinite extension, $\mathfrak{b}X = \#X$, but where X has finite extension, $\mathfrak{b}X = \vartheta^{-1}(\#X)$. Now the anti-cardinalist can hold that 'less than', $<_{\mathfrak{b}}$, is defined by appending the standard order on transfinite cardinals to the standard order on natural numbers; $<_{\mathfrak{b}}$ is now trichotomous over its field. On the basis of this, she can define addition, multiplication, exponentiation, and other operations in a way that is uniform and coherent for natural numbers and transfinite cardinal numbers.

But this version of anti-cardinalism avoids the problems with the ambiguity thesis only by hewing so closely to the important features of cardinalism as to undermine its claims to independent motivation. The structure $(\text{range}(\mathfrak{b}), <_{\mathfrak{b}})$ is precisely isomorphic to $(\mathsf{Card}, <_{Card})$. It is this isomorphism that permits the anti-cardinalist to recapture all our ordinary claims involving canonical names and the 'number of' operator. But now we are confronted with two hypotheses for explaining how our everyday claims about finite and infinite cardinals relate to the mathematical universe: on the first, the 'number of' operator picks out a natural kind and terms for relations and functions have the semantics straightforwardly required for our claims to receive the appropriate truth values; on the second, the 'number of' operator cuts across the boundaries of natural kinds, and terms for relations and functions are gerrymandered in a corresponding way in order to yield the expected truth values. The first hypothesis is undoubtedly simpler and more elegant; in order to adopt the second, we need some independent reason to believe that there is a system of countably many mathematicalia disjoint from Card, forming a meaningful natural mathematical kind, and a reason to believe that canonical numeral terms and expressions of the form ⌜the number of Φs⌝, when there are only finitely many Φs, refer to those objects.

Here the proponent of a cross-kind theory who accepts the $Nat \rightharpoonup Ord$ reduc-

tion—the *ordinalist*—is in a different position from an advocate of the cross-kind theory who holds that natural numbers are *sui generis*. The former has independent reasons to believe in an acceptable collection of objects, reasonably unified in nature, to serve as values of $\flat$: given that the theory T_{Ord} is uniquely interpreted, assertory, and true, there are ordinals and thus finite ordinals; given the failure of the $Card \rightharpoonup Ord$ reduction discussed earlier, they are disjoint from the cardinals. But, in this case, the basic metaphysical picture adopted by the ordinalist is the same as that of the cardinalist: both agree that On and Card exist; neither posits a *sui generis* collection of natural numbers disjoint from both. The debate is merely about the reference of canonical numeral terms and, correlatively, about whether the values of the 'number of' operator lie wholly within Card or are split over Card and On. But since the basic metaphysical posits involved in the two positions are the same, and both positions predict the same linguistic data, all the ordinalist can offer is an unappealing, epicycle-laden alternative to the cardinalist's semantics. The cross-kind theorist who endorses *sui generis* natural numbers, on the other hand, has a metaphysical picture distinct from that of the cardinalist, with an additional category of mathematicalia; if this picture could be supported by independent considerations, the costs involved in the gerrymandered character of $\flat$ might be outweighed. But what could these independent motivations be? There are no distinctive linguistic data, deliverances of general metaphysics, or features of mathematical practice better explained by this account than cardinalism; at most there is the brute intuition that the finite is *special*—so special that it cannot be the case that $0_{Nat}, 1_{Nat}, 2_{Nat}, \ldots$, and $\aleph_0, \aleph_1, \aleph_2, \ldots$, are things of the same fundamental mathematical kind. But this attitude runs counter to the general post-Cantorian development of the field: it elevates the mathematican's merely "medical" (Russell 1936: 143) limitations to an importance unmerited by the deliverances of mathematical science as a whole.

Thus, we have good reason to add $\text{Nat}\, x = \text{Card}\, x \wedge \text{Fin}\, x$ (with an appropriate second-order definition of finitude) and the corresponding definitions of operations to our system. The end result, MB + (GAR) + (NID) + appropriate definitions, is the system Sets, Cardinals, and Ordinals (SCO).

5. Against *Ante Rem* Structuralism

My positive case for the *sui generis* view of cardinals and ordinals has now been set out. At this point, it is worth discussing in detail the differences between my position and the one in the established literature that, in many ways, is closest to it: the *ante rem* structuralism defended by Shapiro (1997).

The distinction is clearest in the case of the natural numbers. I have defended $\mathbb{N} = \{\kappa \in \text{Card} : \kappa < \aleph_0\}$—natural numbers *are* the finite cardinals—and I have offered an account of what it is to be a cardinal: to be one of the values of a special, metaphysically privileged type-$\overline{(e \to t) \to e}$ functor. But I have had little to say about the natural number *structure*—$\mathbb{N}$ with zero and the successor relation distinguished, and perhaps additional distinguished operations. On my view, if by 'natural number structure' we mean an object, then we mean the set $\mathcal{N}_{Ded} = \langle \mathbb{N}, 0, ' \rangle$ or some similar set-theoretical construction, and the general account of its properties is given by set theory: it is not a matter of fundamental concern to the ontology of the natural numbers as individual objects. If we mean 'structure' loosely, then we should ultimately work directly in higher-order terms—and then the important questions will be ones concerning the higher-order structure of reality, and the fineness of grain of Concepts, not the metaphysical relations among objects. In contrast, the *ante rem* structuralist ties the natural numbers very closely to a distinct object, the natural number system. On his view, the natural numbers are places in an A-structure.

In Shapiro's terms, "a natural number is a place in the natural-number structure, a particular infinite pattern. The pattern may be exemplified by many different systems, but it is the same pattern in each case" (1997: 77). Crucially, this 'pattern' is *prior in nature* to the natural numbers themselves:

> Each mathematical object is a place in a particular structure. There is thus a certain priority in the status of mathematical objects. The structure is prior to the mathematical objects it contains, just as any organization is prior to the offices that constitute it. The natural-number structure is prior to 2, just

> as "baseball defense" is prior to "shortstop" and "U.S. Government" is prior to "vice president." (Shapiro 1994: 78)

In other words, in the notation introduced earlier, given any ω-sequence X, the natural numbers make up the A-structure $\$(X)$.[39]

Linnebo has usefully distinguished two independent commitments involved in Shapiro's version of *ante rem* structuralism—the 'incompleteness claim' (discussed above in §1.3) and the 'dependence claim'—the thesis that "mathematical objects from one and the same structure depend on each other and on the structure to which they belong" (2008: 59).

The incompleteness claim is problematic, for reasons discussed in the previous section. The dependence claim is more important, and it represents the crucial point of disagreement between *ante rem* structuralism and the version of *sui generis* antireductionism I defend. Linnebo (2008: 67–68) identifies two distinct components of the dependence claim:

(ODO) For any distinct mathematical objects x, y in the domain of a mathematical A-structure $\mathfrak{S}$, $x \lhd y$.

(ODS) If x is a mathematical object in the domain of a mathematical A-structure $\mathfrak{S}$, then $\mathfrak{S} \lhd x$.

(ODO) clearly violates the asymmetry of dependence; it entails, for instance, that 2_{Nat} depends on 3_{Nat} and vice versa. As such, it conflicts with a basic principle about dependence. But, as Linnebo points out (2008: 69), retreating to the more plausible (ODS) on its own permits the *ante rem* structuralist to preserve most of her basic picture; in particular, it captures something of the notion, often defended by structuralists, of 'relations prior in nature to their relata' (2008: 71).[40]

[39] I use 'ω-sequence' here as a general term for a countable set equipped with added structure, isomorphic to the standard natural numbers, leaving aside the question whether this includes only an ordering or additional operations. There is an additional serious problem for *ante rem* structuralism in the vicinity—the problem of characterizing what is intuitively 'one' structure in various nonequivalent signatures—but I cannot pursue it here.

[40] For examples of this idea, see Ladyman (1998, 2007).

As Linnebo notes, however, (ODS) cannot be a general claim about all mathematical objects, for it conflicts with WBD. A nonempty set a is a paradigmatically simple mathematical structure—a structure with itself as domain—so for every $x \in a$, we have $a \lhd x$ by ODS and $x \lhd a$ by WBD; thus, we have another counterexample to asymmetry. *Ante rem* structuralism about sets is in serious trouble.[41]

Nonetheless Linnebo admits that in some cases (ODS) may hold for particular structures outside set theory. Linnebo's paradigm example is the two-element group (2008: 74), which he plausibly assumes the *ante rem* structuralist will take to be an abstract object, the "abstract form" (2008: 75)—in my terms, the A-structure—shared by every two-element group. Of course, if this is a genuine mathematicale, then $\mathscr{A}_{\mathscr{F}_{Group}}$ will include an assertory, existentially committing R-theory whose quantifiers, on the intended interpretation, range over abstract forms, alongside an algebraic R-theory and a generalist R-theory; in the idiom of §1.2, it will include $\hat{T}_{Group-1}$, $\hat{T}_{Group-2}$, and $\hat{T}_{Group-3}$.[42] Linnebo argues that, assuming there exists such a thing as the abstract two-element group, it obeys (ODS). For my purposes, I can remain neutral on this issue: the more important question here is whether (ODS) holds for Card, On, or $\mathbb{N}$.

(ODS) does not itself conflict with the premisses of the dependence argument, since (D3), (D4), (D3′), and (D4′) say nothing about the dependence of ordinals on non-ordinals, and neither (ODir) nor (CDir) constrains dependence on A-structures. Nonetheless, (ODS) is incompatible with each of the two motivating pictures I proposed—the pure order-type/pure cardinality picture and the bottom-up picture. To be sure, there is a version of the former on which mathematicalia depend on A-structures, but the A-structures in question are those of the mathematicalia themselves (if such things exist), not A-structrures of larger systems within which the mathematicalia are places.

Moreover, there are direct arguments against (ODS) for ordinals, cardinals, and natural numbers. First of all, in none of these cases can (ODS) be formulated in the

[41] A similar point is made in Parsons (1990: 332).

[42] For a similar position about graph theory, see Leitgeb and Ladyman (2008).

most direct way:

(4.5.1) For all $x \in \mathsf{Card}$, $\mathsf{Card} \lhd x$,

(4.5.2) For all $x \in \mathsf{On}$, $\mathsf{On} \lhd x$,

and

(4.5.3) For all $x \in \mathbb{N}$, $\mathbb{N} \lhd x$

are all inadequate. (4.5.1) and (4.5.2) are ill-formed, for On and Card are not genuine singular terms, but shorthand for higher-order locutions, and $\lhd$ stands for a relation between objects; it is not a cross-type operator. Thus, it is crucial for the *ante rem* structuralist that A-structures here be *objects*—things falling under the range of the first-order quantifier; our approach of paraphrasing away expressions like '$\mathscr{S}$' in higher-order terms is unavailable to him.

As for (4.5.3), it directly conflicts with WBD, since $\mathbb{N}$ is simply the set of true natural numbers. We need to differentiate the *set* $\mathbb{N}$ from the objectual A-structure $\mathfrak{N}$ over $\mathbb{N}$ that serves as the 'abstract form' of the natural number sequence, on the *ante rem* structuralist's account: (ODS) yields

(4.5.4) For all $x \in \mathbb{N}$, $\mathfrak{N} \lhd x$,

rather than (4.5.3). But $\mathfrak{N}$, although it must be an object in order for the dependence claims to make sense, cannot be identified with any 'structure' *in the model-theoretic sense* over $\mathbb{N}$, such as the ordered triple $\mathscr{N}_{Ded} = \langle \mathbb{N}, 0, ' \rangle$ or the ordered quintuple $\mathscr{N}_{PA} = \langle \mathbb{N}, 0, ', +, \times \rangle$ that serve as the standard set-theoretical 'intended' models of PA^2 and PA, respectively, in their usual signatures. Since, *e.g.*, $\mathbb{N} \in \mathscr{N}_{Ded}$, by the transitivity of ontological dependence and WBD, we have $x \lhd \mathscr{N}_{Ded}$ for all $x \in \mathbb{N}$, which conflicts with (4.5.4) and asymmetry.

Similarly, if *ante rem* structuralism about cardinals and ordinals is to hold, there will have to exist A-structures $\mathfrak{Card}$ and $\mathfrak{On}$—genuine objects—such that $\mathfrak{Card} \lhd x$ for all $x \in \mathsf{Card}$ and $\mathfrak{On} \lhd x$ for all $x \in \mathsf{On}$. Here, however, a serious problem arises: for

precisely the same reasons that ITA is inconsistent, it cannot be the case that for every proper-class-sized structure (in the higher-order sense) $\mathscr{S}$ there exists some object $\mathfrak{S} = \$(\mathscr{S})$ that is its A-structure; since $\$(\mathscr{S}_1) = \$(\mathscr{S}_2)$ only if $\mathscr{S}_1$ is isomorphic to $\mathscr{S}_2$, this will result in contradiction due to Cantor's theorem, as Linnebo (2008: 75) notes in passing. The crucial point is that an *ante rem* structuralist account of cardinals and ordinals will have to confront the *selectivity problem*—explaining why some higher-order structures have corresponding objectual A-structures and others do not. And the sort of response given by RITA and similar principles—that non-rogue abstracts exist just when the higher-order structures are set-sized—is unavailing here: it will yield abstracts for each individual cardinal and each individual ordinal, but it will not yield $\mathfrak{Card}$ and $\mathfrak{On}$ themselves. For $\mathfrak{N}$, however, this issue does not arise. I shall thus focus on (4.5.4) as the best case for the *ante rem* structuralist; as we shall see, it too has serious drawbacks.

In the case of $\mathfrak{N}$, the *ante rem* structuralist does not confront the best-known objection to his position—the problem of nontrivial automorphisms (Keränen 2001, 2006; Räz 2015; Kouri 2016), which most famously afflicts the *ante rem* structuralist account of the complex numbers and the integers. In contrast to these systems, $\mathfrak{N}$ is *rigid*.[43] Linnebo and Pettigrew (2014: 277) have shown that, for rigid set-sized systems, there is no formal problem in introducing a principle of structural abstraction that fulfils the requirements on \$: where $\mathscr{S}$ is rigid, the principle guarantees the existence of a unique A-structure $\$(\mathscr{S})$ that is both isomorphic to $\mathscr{S}$ and has the property that, for all $\mathscr{S}'$, $\$(\mathscr{S}) = \$(\mathscr{S}')$ if and only if $\mathscr{S}$ is isomorphic to $\mathscr{S}'$. Moreover, they introduce a notion of 'fundamental' property that allows for a weak version of the incompleteness claim: an A-structure has no *fundamental* properties not shared by every structure isomorphic to it.

Nonetheless, the *ante rem* approach to the natural numbers faces a different problem, which generalizes to other putative A-structures: the problem of *sub-* and *super-*

[43] The best-known case is that of the complex numbers, since $a+bi \mapsto a-bi$ is a nontrivial automorphism on $\mathbb{C}$. In any case, arguments that *ante rem* structrualism fails for structures with nontrivial automorphism rely on strong (and contestable) versions of the incompleteness claim.

structure relationships. Recall that $\mathfrak{N}$ itself is an ω-sequence; thus, even though it is not a set-theoretical construction, it must have some form of internal complexity: only if this is the case can it make sense to speak of natural numbers as 'places' in $\mathfrak{N}$. But if this is the case, then $\mathfrak{N}$ presumably has substructures; in particular, for every $n \in \mathbb{N}$, there exists a structure $\mathfrak{N}^{\restriction n}$ with domain $\{m \in \mathbb{N} : m < n\}$ and the appropriate restrictions of all the relevant operations on $\mathfrak{N}$. (We can leave open whether the only relevant operation is successor, as in $\mathscr{N}_{Ded}$, or whether others are in some sense 'incorporated' into $\mathfrak{N}$.) But here a dilemma arises: is $\mathfrak{N}^{\restriction n}$ (holding n fixed as a parameter) itself an A-structure?

Call any structure isomorphic to $\mathfrak{N}^{\restriction n}$ an *n-sequence*; if $\mathfrak{N}^{\restriction n}$ is an A-structure, then for any n-sequence $\mathscr{S}$, we have $\$(\mathscr{S}) = \mathfrak{N}^{\restriction n}$. But if this is the case, then by (ODS), $\mathfrak{N}^{\restriction n} \lhd m$ for every $m, n \in \mathbb{N}$ where $m \leq n$. This is itself already intuitively unappealing, for it means that (for example) 4_{Nat} depends not only on $\mathfrak{N}$, but on $\mathfrak{N}^{\restriction 4}$, $\mathfrak{N}^{\restriction 5}$, and so on. Moreover, plausibly $\mathfrak{N}^{\restriction n} \lhd \mathfrak{N}$ for every n, on the basis of intuitions similar to those which motivate WBD; thus, by the transitivity of ontological dependence, 4_{Nat} depends on $\mathfrak{N}$ in multiple ways: mediately by one path, immediately by another. This two-path dependence seems to violate the *ante rem* structuralist's principle that natural numbers are essentially characterized *directly* by their place in $\mathfrak{N}$; instead, a more complex system of relationships is required.

But the other horn of this dilemma is also unappealing: if, for instance, $\mathfrak{N}^{\restriction 4}$ is not an A-structure, then there is a different A-structure, $\mathscr{N}_4 = \$(\mathfrak{N}^{\restriction 4})$, which characterizes its abstract form. It might be objected that this consequence is unacceptable, for something that figures in any A-structure ought to be essentially the kind of thing that can only appear in A-structures. This principle, however, is one that the *ante rem* structuralist will have to reject, because of the permutation problem raised by Hellman (2001: 195–96): any nontrivial permutation π on $\mathbb{N}$ will induce a structure $\mathfrak{N}^{[\pi]} \neq \mathfrak{N}$ in which the natural numbers form an ω-sequence, although not in their standard order; yet $\mathfrak{N}^{[\pi]}$ cannot be an A-structure, since $\$(\mathfrak{N}^{[\pi]}) = \mathfrak{N}$. In the general case, the permutation problem is probably not a decisive objection to *ante rem* structuralism, at least

for rigid A-structures: the *ante rem* structuralist can plausibly claim that it is essential to o_{Nat} that it be *in initial position* whenever it figures in the A-structure of the natural numbers; when it occupies a place in a structure that fails to meet this requirement, it is not surprising that the result should be something other than an A-structure. But in the case of *substructures*, this line of response is less appealing: $\mathfrak{N}^{\restriction 4}$ is not generated from $\mathfrak{N}$ by permutation, but by mere truncation. The non-Cambridge properties of all the objects in the domain of $\mathfrak{N}^{\restriction 4}$ relative to one another remain as they are in $\mathfrak{N}$: it is hard to see what more is required for $\mathfrak{N}^{\restriction 4}$ to be $\mathscr{N}_4$.

An even more serious problem is posed by superstructures. In some cases, when an A-structure $\mathfrak{S}$ figures as a position in another structure $\mathfrak{A}$, it is reasonable to claim that $\mathfrak{A}$ is not itself an A-structure. For instance, the symmetric group on $\mathfrak{N}$ is plausibly not an A-structure, because none of the $\mathfrak{N}$-relevant properties of the natural numbers figure in any way in its group-theoretic structure: $\mathbb{N}$ here serves merely as a countable set. But we can conceive of examples where the structural properties of $\mathfrak{S}$ are 'carried along' into the superstructure. Consider, for example, the class of ω-sequences, each carrying all associated structure (i.e., whatever operations are present on $\mathfrak{N}$): we denote this by $\widetilde{\mathfrak{N}}$. Let $\mathfrak{Omega}$ be the category with objects given by $\widetilde{\mathfrak{N}}$, arrows given by the class of bijections on ω-sequences, and composition of arrows defined as composition of the underlying bijections. Plausibly, $\mathfrak{Omega}$ is itself an A-structure. (Of course, it is a large category, and thus a class-sized structure, so the same problem of selectivity arises that is present with $\mathfrak{Card}$ and $\mathfrak{On}$; for the purposes of the illustration, however, we can leave this aside.) But then $\mathfrak{Omega} \lhd \mathfrak{N}$, and by the transitivity of dependence, $\mathfrak{Omega} \lhd n$ for each $n \in \mathbb{N}$. Yet this does not comport well with the basic *ante rem* structuralist idea: the natural numbers now essentially depend on structures *outside* $\mathfrak{N}$, as well as on $\mathfrak{N}$ itself.

These issues arise in large measure because the *ante rem* structuralist is committed not only to the truth of abstraction principles that generate abstracts for all structures (in the higher-order sense), but to a controversial account of dependence relations between the corresponding abstracts (A-structures) and other mathematicalia. The for-

mer is unproblematic: indeed, it is very similar to the picture behind RITA^n (although the idea of structure abstraction, by conflating relations of different adicity, fails to respect appropriate distinctions of type structure). The latter, however, has no justification. The RITA^n family can do the work of the more appealing components of the *ante rem* structuralist's view, without its drawbacks.

V: DEPENDENCE, ESSENCE, AND GROUNDING

In the previous chapter, I argued for (D3′)—the thesis that no cardinal depends on any cardinal—informally, on the basis of the pure-cardinality type conception. What it is to be the cardinal κ is to be the representative of those Concepts with κ-membered extensions; the only way that κ could depend for its being on an object would be if that object were in the extension of *every* κ-membered class, for only thus could its existence be indispensable to the existence of κ. This idea justifies not only (D3′) but the stronger cardinal independence thesis,

(CInd) $\quad \forall\kappa\neg\exists x(x \lhd \kappa)$,

although this is not required for the argument in §4.1.

It is worth asking, however, whether it is possible to render this chain of reasoning more explicit by appealing to bridge theses that explicitly link abstraction principles and dependence. There are two plausible ways to do this: the first uses the notion of *essence*, the second the notion of *grounding*.[1] Gideon Rosen and Stephen Yablo (MS) have pursued the former, Robert Schwartzkopff (2011) the latter (although the notion of grounding also figures, in passing, in Rosen and Yablo's paper). Rosen and Yablo are primarily interested in the Julius Caesar problem, not in objectual ontological dependence; Schwartzkopff is interested in dependence, but he draws a conclusion incompatible with (CInd): he argues that a proper understanding of the interaction between grounding and HP yields the result that every cardinal except o_{Card} depends on all its predecessors. In contrast, I maintain that there is a plausible argument for (CInd) from a grounding-theoretic version of HP, GHP. Thomas Donaldson (MS) terms a variant on GHP the 'Schwartzkopff–Rosen Principle' and argues that, given standard assumptions about the behaviour of quantifiers in grounding contexts, it leads to unacceptably promiscuous dependence relations. I show that Donaldson's challenge can

[1] For survey articles on the former, see Clark and Liggins (2012), Correia and Schnieder (2012b), Trogdon (2013), and Raven (2015). Key papers on grounding include Schaffer (2010), Rosen (2010), Fine (2010a, 2012a, 2012b, 2015), Batchelor (2010), Correia (2014), Schnieder (2011), and Audi (2012a, 2012b). On essence, see Fine (1994a, 1994b, 1995a, 1995b), Koslicki (2012), and Correia (2005).

be met, given a proper understanding of the nature of the equinumerosity operator in GHP.

5.1 Essence

One obvious approach to formulating the link between abstraction principles and dependence involves a formalized notion of essence, which Kit Fine (1994a, 1994b, 1995a) has treated extensively. Essence claims are intended to capture distinctions among ways in which an object can possess certain properties, or ways in which truths about an object or property can obtain, that are finer-grained than those which can be captured with modal sentential operators alone. The example that Fine uses to motivate the distinction is the same as the one he uses to introduce objectual dependence:

> Consider, then, Socrates and the set whose sole member is Socrates. It is then necessary, according to standard views within modal set theory, that Socrates belongs to singleton Socrates if he exists; for, necessarily, the singleton exists if Socrates exists, and, necessarily, Socrates belongs to singleton Socrates if both Socrates and the singleton exist. It therefore follows according to the modal criterion that Socrates essentially belongs to singelton Socrates.
>
> But, intuitively, this is not so. It is no part of the essence of Socrates to belong to the singleton. Strange as the literature on personal identity may be, it has never been suggested that in order to understand the nature of a person one must know to which sets he belongs. (1994a: 4)

This leads to a first concept of essence: *objectual essence*, which is represented using a type-$(t, e \to t)$ operator—a sentential operator with an index for a singular term. Following Fine, we write $\Box_x$ for the operator, which can be glossed as 'it holds in virtue of the essence of x that'. Fine's example thus amounts to the conjunction of the theses

$$\Box_{\{\text{Socrates}\}}\text{Socrates} \in \{\text{Socrates}\} \tag{5.1.1}$$

and

(5.1.2) $$\neg\Box_{\text{Socrates}}\text{Socrates} \in \{\text{Socrates}\}.$$

In addition to objectual essence operators, Fine introduces apparatus for what might be called *Conceptual essence*, which we can represent as a family of operators of type $(t,(e \to t) \to t)$, type $(t,(e,e \to t) \to t)$, type $(t,(e,e,e \to t) \to t)$, and so on—sentential operators $\Box_{X^n}$, with an index for an n-ary predicate term. We can think of these as 'it holds in virtue of the essence of being X' (or, perhaps more idiomatically, 'in virtue of what it is to be X^n') operators. Thus, for example, it holds *in virtue of the essence of being a sibling* that if x and y are siblings, then they share at least one parent:

(5.1.3) $\Box_{(\lambda xy.x \text{ is a sibling of } y)}\forall x\forall y\,(x \text{ is a sibling of } y \to \exists z(z \text{ is a parent of } x \wedge z \text{ is a parent of } y))$.

Fine suggests that we can define objectual essence in terms of Conceptual essence as follows,

(5.1.4) $$\Box_x\phi \stackrel{\text{Df.}}{=} \Box_{(\lambda z.z=x)}\phi,$$

but for our purposes we can be neutral on this question; informally, we shall use both notions.

One way of understanding the connexion between essence and dependence is to imagine that the sentences figuring inside an essence operator denote structured Russellian propositions. This is, of course, incompatible with our primitivist type-theoretical approach, but it can serve as a heuristic to motivate principles that can then be reformulated more cleanly in the official idiom. Let us write $\langle\phi\rangle$ for the structured proposition corresponding to ϕ: if $\Box_a\phi$ and b is a constitutent of $\langle\phi\rangle$, then, according to Fine, a depends on b. Moreover, we can think of an object a as characterized by a collection of *essence facts* $\langle\phi_1\rangle,\langle\phi_2\rangle,\ldots$, corresponding to true essence claims of the form $\Box_a\phi_1,\Box_a\phi_2,\ldots$ (although, in general, they will not be capable of being expressed in a

finite language). Then, assuming that structured propositions of the relevant type are closed under conjunction, the *total essence* of a will be $\langle\vec{\phi}\rangle = \langle\phi_1 \wedge \phi_2 \wedge \ldots\rangle$. The total essence provides something like a real definition of a; on the Finean picture, a depends objectually on an object b if and only if b is distinct from a and is a constituent of a's total essence.[2] We can then introduce *total essence operators* $\blacksquare_x$ and $\blacksquare_{X^n}$ ('it holds in virtue of the total essence of x that', 'it holds in virtue of the total essence of being X^n that'), with the essence–dependence link represented as

(EssDep) $\quad \forall x \forall y (x \lhd y \leftrightarrow x \neq y \wedge x$ is a constituent of the unique p such that $\blacksquare_y p)$.

How can this apparatus be applied to the question of constitutiveness for cardinals? One natural thought is that it is part of the essence of every cardinal, or of being a cardinal, that HP obtains. But neither of the obvious formulations of this claim,

$$\Box_{\text{Card}} \forall X \forall Y (\#X = \#Y \leftrightarrow X \approx Y) \tag{5.1.5}$$

and

$$\forall x (\text{Card } x \rightarrow \Box_x \forall X \forall Y (\#X = \#Y \leftrightarrow X \approx Y)), \tag{5.1.6}$$

is particularly useful as a basis for a dependence thesis. (Here we use the Card as defined in §4.3). Even (5.1.6) fails to single out any essence fact about any *particular* cardinal; the variable x appears only in the index, not the prejacent. Thus (5.1.5–6) cannot be the whole story about the essence of cardinals.

What is needed, it seems, is to quantify into the essence context to get individualized claims about particular cardinals. There are technical difficulties involved here, because essence contexts are non-congruential: substitution of logical equivalents *salva veritate* is not generally available inside an essence operator. Nonetheless, although

[2] This is slightly different from Fine's official formulation (1995b: 275) because of Fine's appeal to the equivalence in (5.1.4). See generally Koslicki (2012: 196): "There is not, on this approach, much of a difference between essences and real definitions." The distinctness requirement is not given by Fine or Koslicki, but it is clearly necessary, since, for instance, $\Box_a a = a$.

essence is hyperintensional in this sense, we have no reason to think that it is hyperintensional in the way that cases of Frege's puzzle are: we can assume that substitution of co-referring proper names (and, likewise, rigid designators for higher-order entities, including higher-order variables) preserves truth. (In view of this assumption, we could represent $\Box_x$ formally using M-models, although we shall not attempt to do so.)

If we confine ourselves to the natural numbers (finite cardinals), however, we can remain at the level of schemata. Let us assume that '0_{Card}', '1_{Card}', and so on are directly referential and thus available for use in specifying essence facts without incurring covert additional commitments. We can then posit that the constitutive relationship between the cardinals and HP yields a series of essence claims of the form:

$$\Box_{0_{Card}} \forall X (\#X = 0_{Card} \leftrightarrow \ldots),$$
$$\Box_{1_{Card}} \forall X (\#X = 1_{Card} \leftrightarrow \ldots),$$
$$\Box_{2_{Card}} \forall X (\#X = 2_{Card} \leftrightarrow \ldots),$$
$$\ldots,$$

with the ellipsis filled in based on the details of some account extracted from the general form of HP. One such account is provided by Frege's own account of the natural number sequence in §§74–78 of *Grundlagen*.[3] If we treat the terms used in the construction as, in effect, real definitions, we obtain

$$\Box_{0_{Card}} \forall X (\#X = 0_{Card} \leftrightarrow X \approx (\lambda x \,.\, x \neq x)),$$
$$\Box_{1_{Card}} \forall X (\#X = 1_{Card} \leftrightarrow X \approx (\lambda x \,.\, x = 0_{Card})),$$
$$\Box_{2_{Card}} \forall X (\#X = 2_{Card} \leftrightarrow X \approx (\lambda x \,.\, x = 0_{Card} \vee x = 1_{Card})),$$
$$\ldots.$$

This set of claims has the result of making 1_{Card} depend objectually on 0_{Card}, 2_{Card} depend on 1_{Card}, and so on—contradicting both the strong and the weak independence

[3] I thank Kit Fine for pressing me on this point.

theses.

This is, as we shall see, essentially the approach taken by Schwartzkopff in his discussion of HP and grounding. But there is a problem with viewing this way of introducing terms for the natural numbers as generating real definitions, which Frege himself noted: instead of $(\lambda x \,.\, x \neq x)$, we "could have taken any other Concept under which nothing falls": non-self-identity is merely "the most convenient" (Frege 1884: §74/1980: 86).[4] For instance, we could choose to introduce the term '1_{Card}' by way of the definition $\#(\lambda x \,.\, x = \infty_{\#})$, having previously introduced '$\infty_{\#}$' by way of the definition $\#(\lambda x \,.\, x = x)$. If this way of picking out 1_{Card} corresponded to a real definition, however, we would have $\Box_{1_{Card}} \forall X(\#X = 1_{Card} \leftrightarrow X \approx (\lambda x.x = \infty_{\#}))$, with the perverse result that the 1_{Card} depends objectually on anti-zero; similarly, we can easily devise circular collections of putative real definitions.

What has gone wrong here is that a fundamental fact about the *generality* of HP has been missed: the number 3_{Card} is constitutively an abstract in virtue of its relationship not to some particular Concept under which three things fall, but to *every* Concept under which three things fall. This is the idea that Rosen and Yablo (MS) develop. One of the many ways in which HP is explanatorily rich is that it has its application conditions 'built in': any adequate definition of natural number terms will deliver the *Frege equivalence schema*,

$$\forall X(\#X = \dot{n} \leftrightarrow \exists_{=n} x\, X x), \tag{5.1.7}$$

where $\dot{n}$ is a canonical term for n and $\exists_{=n}$ is the standard numerical quantifier.

Rosen and Yablo propose to take these equivalences as 'exhaustively definitive' of the finite cardinals. Their exhaustive definition operator $\star_x$ is defined as the conjunction of the $\vartriangle_x$ ('it is definitive of a that … ') and $\triangledown_x$ ('nothing beyond … is definitive of a') operators (MS: 8). Since they are explicit that *real* definitions are at issue, the

[4] "Ich hätte zur Definition der 0 jeden anderen Begriff nehmen können, unter den nichts fällt. Es kam mir aber darauf an, einen solchen zu wählen, von dem dies rein logisch bewiesen werden kann; und dazu bietet sich am bequemsten »sich selbst ungleich« dar […]."

intended meaning of the operator $\star_x$ will be fundamentally the same as that of the $\blacksquare_x$ operator introduced above.[5] Rosen and Yablo argue:

> Suppose that there exist *Frege numbers.* A Frege number is a thing with a distinctive sort of essence. For each Frege number n there is a cardinality quantifier, $\exists_n$, such that
>
> $\star_n$ For all F, $(n = \#F \leftrightarrow \exists_n x\, Fx)$.
>
> So, for instance, the Frege number 2 is an item that is, by definition, the number that numbers the Fs iff there is an F, x, and another F, y, and every F is either x or y. Also, though, and just as important, *all* essential facts about the Frege number 2 are implicit in the one just mentioned. This means that it is not essential to this item that, say, there are or could be human beings (MS: 16).

Rosen and Yablo argue that this rules out 2_{Card} = Julius Caesar; if the argument is successful, however, it also rules out the dependence of any finite cardinal on any object unless that dependence is explicable through the quantifier and second-order variable, since those are the only expressions figuring in the statement of the total essence that could correspond to a constituent or some constituents. But—returning for a moment to the Russellian heuristic—it seems clear that a *bound* second-order variable cannot represent any constituent of a structured proposition, just as a bound first-order variable cannot. Likewise, nothing can be a constituent of $\langle \forall X\, (\dot{n} = \#X \leftrightarrow \exists_n x Xx)\rangle$—which we shall term the *Frege fact for n*—merely in virtue of being a potential value of the bound second-order variable. And since the other terms in the sentence expressing the Frege fact—the identity predicate and logical constants—also represent nothing capable of carrying commitment to any particular objects, no object (except the cardinal number n itself) is a constituent of the total essence of n. From this it follows,

[5] Rosen and Yablo use a star of David (*i.e.*, an overlapping $\triangle$ and $\triangledown$) for the operator. Typographical limitations have caused us to substitute $\star$.

by (EssDep), that n depends on nothing: thus (CInd) holds for finite cardinals.[6]

Unfortunately, the claim that the Frege fact for n *exhausts* the total essence of n is not plausible, if exhaustion is understood in the natural way: *every* fact about what is essential to n must be logically entailed by the Frege fact for n. A problem arises similar to that faced by the *ante rem* structuralist committed to the 'incompleteness claim', discussed in §1.3 (following Linnebo 2008). 26_{Card} exists necessarily; this does not follow from its Frege fact. If necessitism is true, this may not be a serious problem: if everything exists necessarily, then there is no reason to think that any one particular object's existing necessarily should figure in its essence. (This can be compared with the treatment of logical truths in Fine's (1995a) logic of essence. The essence operator is not, in general, closed under logical consequence, for not all logical truths are relevant to the essence of an object.) But if contingentism is true, then it is quite an implausible consequence. In any case, other non-mathematical properties will pose problems: 26_{Card} is necessarily abstract, necessarily non-spatiotemporal, and so on. The claim that, for each cardinal n, these facts about it hold in terms of its essence, is intuitively compelling—but the Frege fact for n, or all the Frege facts together, cannot yield it. Moreover, although all the mathematical properties of the cardinals will be entailed by all the Frege facts taken together (along with a powerful enough background logic), mathematical properties will not be localized properly. That 17_{Card} is prime should lie in *its* essence. But it is not entailed by $\forall X(17_{Card} = \#X \leftrightarrow \exists_{17} x\, Xx)$, even given the appropriate definitions, without the Frege facts for the other cardinals.

This illustrates a general problem for any attempt to characterize the constitutiveness–dependence link by way of essence: if robust realism about mathematicalia is true, then mathematicalia, like other objects, will have a great many essential properties, mathematical and non-mathematical. On that assumption, (EssDep) may be true, but it is not very useful, since it involves intractably complex, conjunctive total essence facts. This does not mean that the task of delimiting the dependence profile of math-

[6] Presumably the argument can be extended to infinite cardinals, using an infinitary language with canonical names, but this raises difficult issues.

ematicalia is hopeless: on the basis of an independent grasp of the dependence facts, or in some other way, we may know enough to know that no objects are constituents in the total essence of a mathematicale without being able to formulate a sentence expressing that total essence. But if this is the case, then we have simply returned to the rough and ready approach to justifying (CInd) with which we began.

5.2 Grounding

A different approach draws on the idea, adumbrated in §3.1, of *internal explanation*: the right-hand side of an abstraction principle like HP *explains* the left-hand side in a way that (according to the proponent of AAR) reveals metaphysical structure, not mere anthropocentric preferences. This intuition that there are such cases of genuine metaphysical explanation lies behind the theory of *grounding* or *metaphysical propositional dependence*.[7] The governing idea is to regiment those constructions in natural language which plausibly reflect metaphysical explanation—for instance, the non-causal 'because' in sentences such as

(5.2.1) Napoleon was English or French because he was French,

(5.2.2) The silk is red in virtue of being vermillion,

(5.2.3) There exists an even prime because two is an even prime.

There are also philosophically more contentious examples, where appeals to grounding are made in order to strengthen a thesis that modal notions such as supervenience fail to capture fully:

(5.2.4) James is in pain because his C-fibres are firing,

(5.2.5) This action is bad because it fails to maximize the total intrinsic good of the universe.

Of course, those who reject these accounts of mental states or ethical properties will deny that (5.2.4–5) are true; nevertheless, the sense of 'because' that the nonreductive

[7] My presentation of the general ideas behind grounding follows Rosen (2010) and Fine (2012b).

physicalist or Moorean consequentialist who proposes these theses is aiming at should be clear.

The easiest way to formulate these claims is usually with the locution ⌜the fact that ϕ obtains in virtue of the fact that ψ⌝ (Rosen 2010). But it is more elegant to develop an account of grounding in 'operational' rather than 'predicational' terms (Correia 2010: 253), while continuing to use the idiom of facts informally. Most of the philosophers who have taken this tack have treated grounding as something like an object-language version of a sequent connective, taking a set or plurality of sentences (displayed as a list) on the left and a single sentence on the right. Thus, for example, (5.2.1) is regimented as

(5.2.6) Napoleon was French $\prec$ Napoleon was English $\vee$ Napoleon was French,

with $\prec$ as the grounding connective. Since Napoleon was French, (5.2.6) follows from the schemata

$$\phi \rightarrow (\phi \prec \phi \vee \psi), \tag{5.2.7}$$

$$\psi \rightarrow (\psi \prec \phi \vee \psi), \tag{5.2.8}$$

expressing the thesis that a true disjunction is grounded in any true disjunct. (We adopt the convention that $\prec$ binds more loosely than other connectives.) Likewise, a true conjunction is grounded *jointly* in both its true conjuncts:

$$\phi \wedge \psi \rightarrow (\phi, \psi \prec \phi \wedge \psi). \tag{5.2.9}$$

These are relatively uncontroversial principles about grounding, as is the claim that a true existential quantification is fully grounded in each witnessing instance,

$$\phi(a) \rightarrow (\phi(a) \prec \exists x\, \phi(x)). \tag{5.2.10}$$

The same is true of some basic structural principles. We let Γ, Δ, and so on stand for pluralities of sentences. Grounding is factive: if $\Gamma \lessdot \phi$, then ϕ obtains and so does every $\psi \in \Gamma$. It also has the natural generalizations of transitivity and antisymmetry—the *cut* and *noncircularity properties*.[8] When there is only one expression on the left of the grounding connective, these reduce simply to operational versions of transitivity and antisymmetry.

There are some problems with this approach, however: $\lessdot$ is treated as syncategorematic here, which conflicts with the principle that every term other than λ should have a type. Furthermore, it is not obvious how to move beyond schematic generality to permit full quantification into grounding contexts, since the left-plural form of the operator is not a standard logical device.

This problem suggests an even more elegant approach: if we take the grounding connective to be a type-$((t \to t), t \to t)$ operator, we can replace the plurality of propositions on the left with a higher-order propositional operator (*i.e.*, in effect, a property of propositions), and integrate grounding directly into the type structure. So the official form of a claim like (5.2.9) will be

$$\forall p \forall q (p \wedge q \to (\lambda p' \mathbf{.}\, p' \equiv p \vee p' \equiv q) \lessdot p \wedge q). \tag{5.2.11}$$

In general, however, we shall continue to write grounding claims in the form of (5.2.7–10), treating this as an abbreviation for the type-$((t \to t), t \to t)$ operator idiom. For convenience, however, we shall introduce *pluripropositional quantification* (represented by pp, qq) to allow us to quantify into the left-hand side of a grounding expression: in actuality, of course, these are type-$(t \to t)$ variables in disguise. We then define

[8] In rule form, as given in Fine (2012b: 22), they are $\dfrac{\Gamma_1 \lessdot \phi_1; \ldots; \Gamma_n \lessdot \phi_n; \phi_1, \ldots, \phi_n \lessdot \psi}{\Gamma_1 \cup \cdots \cup \Gamma_n \lessdot \psi}$ and $\dfrac{\Gamma \cup \phi \lessdot \phi}{\bot}$. Cf. Rosen (2010: 115–16). We have no need of Fine's controversial amalgamation rule (2012b: 57), $\dfrac{\Gamma_1 \lessdot \phi; \ldots; \Gamma_n \lessdot \phi}{\Gamma_1 \cup \cdots \cup \Gamma_n \lessdot \phi}$.

partial grounding,

$$p \dot{\ll} q \stackrel{\text{Df.}}{=} \exists pp(p \prec pp \wedge (pp \lessdot q)), \tag{5.2.12}$$

and, more importantly for our purposes, *ultimate grounding*—grounding by a collection of facts none of which itself has any further ground:

$$pp \lessdot_U q \stackrel{\text{Df.}}{=} pp \lessdot q \wedge \forall p(p \prec pp \rightarrow \neg \exists qq(qq \lessdot p)). \tag{5.2.13}$$

(Here '$\prec$' is to be read as 'is one of', as in standard plural logic.)

Is there a link between grounding and dependence corresponding to (EssDep)? One initially plausible candidate is the thesis that an object a depends on an object b just in case b is distinct from a and a constituent of some fact in every ultimate ground for the fact that a exists:

$$\text{(GrDep1)} \quad \forall x \forall y(x \lhd y \leftrightarrow x \neq y \wedge (\forall pp((pp \lessdot_U (\exists z\, z = y)) \rightarrow \exists p(p \prec pp \wedge x \text{ is a constituent of } p))).$$

Like (EssDep), (GrDep1) is only semi-formal: 'constituent' is motivated by the heuristic Russellian picture, and would have to be replaced by something more rigorous, in keeping with the general type-theoretical treatment, in a full presentation. (There is of course no problem with introducing a $(e, t \rightarrow t)$-type 'constituent' operator, but it cannot be explained by a notion of literal parthood, for on the primitivist picture, 'propositions' are not objects and thus cannot have objects as parts.) Unfortunately, however, (GrDep1) faces counterexamples. Consider the case of {{Napoleon}}. Plausibly, $\exists x(x = \{\text{Napoleon}\}) \lessdot \exists x(x = \{\{\text{Napoleon}\}\})$ and $\exists x(x = \text{Napoleon}) \lessdot \exists x(x = \{\text{Napoleon}\})$ while $\exists x(x = \{\text{Napoleon}\})$ itself has no ground; thus, $\exists x(x = \text{Napoleon}) \lessdot_U \exists x(x = \{\{\text{Napoleon}\}\})$. But this yields only Napoleon $\lhd$ {{Napoleon}}, whereas we also require Napoleon $\lhd$ {Napoleon}.

A general solution to this problem will be difficult to find, but in the cases we are interested in a simplification suggests itself. In general, it will not be possible to define grounding *simpliciter* on the basis of *immediate* grounding, for reasons similar to those which prevent us from defining dependence simpliciter in terms of immediate dependence. In finite cases, however, a definition using the ancestral is efficacious. And it is implausible that any infinite grounding chains will be involved in very simple existence claims of the type that concern us here. We thus temporarily introduce a new primitive, $\lessdot_I$, for *immediate* grounding, and we posit

(GrDep2) $\forall x \forall y (x \lhd_I y \leftrightarrow (x \neq y \wedge \forall pp((pp \lessdot_I (\exists z\, z = y)) \rightarrow \exists p(p \prec pp \wedge x \text{ is a constituent of } p)))$;

then, for the restricted cases under consideration, we assimilate $\lhd$ to $\lhd_I^*$.

The advantage of (GrDep2) over (EssDep) is that (GrDep2) focusses directly on metaphysical explanations for existence. In contrast, facts that obtain in virtue of what kind of thing something is—such as the fact that 3_{Card} is necessarily non-spatiotemporal—count as part of 3_{Card}'s total essence, but they do no real work in explaining its existence. Now we are in a position to mount an argument for (CInd) by using grounding to capture the asymmetrical internal explanatory force of HP. The argument starts from a ground-theoretic version of HP:

(GHP) $\forall X \forall Y (X \approx Y \rightarrow (X \approx Y \lessdot_I \#X = \#Y))$.

One instance of GHP is

(5.2.14) $\forall X (X \approx X \rightarrow (X \approx X \lessdot_I \#X = \#X))$,

which trivially yields

$$\forall X(X \approx X \leqslant_I \#X = \#X). \tag{5.2.15}$$

But by (a universally quantified analogue of) (5.2.10) for immediate dependence, we have $\forall x([x = x] \leqslant_I [\exists y\, y = x])$. Then, using the cut property, we obtain

$$\forall X([X \approx X] \leqslant [\exists x(\#X = x)]). \tag{5.2.16}$$

For every Concept F, the trivial logical fact that there exists a permutation of the things that are F grounds the existence of the object that is the cardinal number corresponding to F. (I am assuming here that 'being identical to something' is the correct way to formulate an existence claim for an object. But even if we had used a primitive existence predicate $E!$ instead, on any plausible account we should have either $\forall x(x = x \leqslant_I E!x)$ or $\forall x(E!x \leqslant_I x = x)$. In the former case, the argument to (5.2.16) would go through; in the latter, anything that grounds $a = a$ at all will have to stand below $E!x$ in a grounding chain. In either case, an analogue of the argument presented here would go through.)

Furthermore, it is reasonable to think that, for every Concept F, $F \approx F$ is itself ungrounded. At the very least, it is difficult to see what could ground it. So (5.2.16) can be strengthened to

$$\forall X([X \approx X] \leqslant [\exists x(\#X = x)]). \tag{5.2.17}$$

But the dependence in (5.2.17) is mediated only by $\#X = \#X$, and it is plausible that this should not introduce any new elements into the dependency chain, as {Napoleon} did. Thus, applying (GrDep2) twice and taking the ancestral, we have the result that, if a cardinal κ depends on an object a, then a will have to be a constituent of $F \approx F$ for every Concept F such that $\#F = a$. But surely nothing can be a constituent of $F \approx F$ for a given Concept F unless it falls within the extension of F. And, of course, nothing falls within the extension of every κ-membered concept, yielding (CInd).

5.3 Schwartzkopff's Argument and Donaldson's Objection

GHP does not figure explicitly in Rosen and Yablo's argument, although Donaldson (MS: 8) correctly notes that it is at least a background assumption of their paper. Schwartzkopff, however, explicitly endorses its necessitation, which he terms 'First-Level Explanatory Hume's Principle' (2011: 362) (although he uses a formalized 'because', which we can take to be equivalent to $\lessdot$ when Γ in $\Gamma \lessdot \phi$ is a singleton).

He further endorses a dependence thesis similar to ours:

(5.3.1) x depends on $yy \stackrel{\text{Df.}}{=} \exists F \Box(\exists z(z = x) \rightarrow \exists z(z = x)$ because Fyy.

Thence, however, he obtains a very different conclusion from ours: "every natural number greater than 0 is an ontological [*sic*] dependent object because it depends for its existence on its predecessors" (2011: 369 fn. 24).

Schwartzkopff's argument is complex, and I shall not spell out every step, but the origin of the incompatibility in our conclusions is straightforward: he "(schematically) define[s]—echoing Frege's own definitions [...]—that (Def. n): for every natural number n greater than 0: $x = n \equiv_{\text{def}} x = N[\text{the } xx : x = 0 \vee \cdots \vee x = n - 1]$" (2011: 267). He terms these "canonical N-terms", and only admits canonical N-terms as substituends for a numeral n in a grounding context. Thus (returning to our notation) Schwartzkopff would accept $(\lambda x \,.\, x \text{ is a region of Gaul}) \approx (\lambda x \,.\, x \text{ is a region of Gaul}) \lessdot \exists x(x = \#(\lambda x.x \text{ is a region of Gaul}))$ as an instance of (5.2.16), but would not accept the transition thence to $(\lambda x \,.\, x \text{ is a region of Gaul}) \approx (\lambda x \,.\, x \text{ is a region of Gaul}) \lessdot \exists x(x = 3_{Card})$ as a permissible substitution, for he denies that intersubstitutivity *salva veritate* holds for '$\#(\lambda x.x$ is a region of Gaul)' and '3_{Card}' in a grounding context; the only permissible substituend for '3_{Card}', on Schwartzkopff's view, would be '$(\lambda x.x = 0_{Card} \vee x = 1_{Card} \vee x = 2_{Card})$', for only that is a canonical N-term. My response to this has already been given in §5.1: it is an error to think that Frege's definitions are in any way *metaphysically* privileged. More fundamentally, we should not think of terms in a canonical notation system such as '0_{Card}', '1_{Card}', and so on as descriptive at all. Although they are *introduced* by definitions such as '$0_{Card} = \#(\lambda x \,.\, x \neq x)$', they are incorporated into

the language as directly referential terms: we could perhaps best think of '0_{Card}' as 'dthat: the number of non-self-identical objects' (Kaplan 1978).

In contrast to Schwartzkopff, Donaldson (MS) maintains that his variant on GHP, in unmodified form, is objectionable because it leads to a plethora of unintended grounding facts. It is important to point out that the principle that he uses is a *variant* of GHP, not GHP itself; nor is the difference confined merely to the use of a predicational idiom for grounding claims. His Schwartzkopff–Rosen Principle, in operational notation, is:

(SRP) $$\forall X \forall Y (\#X = \#Y \rightarrow (\exists W\, X \sim_W Y \lessdot \#X = \#Y)),$$

where $X \sim_W Y$ expresses that W bijects X and Y. (W is here a type-$(e, e \rightarrow t)$ variable.) Moreover, he is explicit that $\sim_W$ is not to be taken as primitive, but as a metalinguistic abbreviation:

> More formally, $X \sim_W Y$ abbreviates the conjunction of the following four conditions:
>
> $\text{Function}_{X,Y}(W)$: $\forall x(Xx \rightarrow \forall y_1(Yy_1 \rightarrow \forall y_2(Yy_2 \rightarrow (Wxy_1 \wedge Wxy_2 \rightarrow y_1 = y_2))))$
>
> $\text{Injective}_{X,Y}(W)$: $\forall y(Yy \rightarrow \forall x_1(Xx_1 \rightarrow \forall x_2(Xx_2 \rightarrow (Wx_1y \wedge Wx_2y \rightarrow x_1 = x_2))))$
>
> $\text{Total}_{X,Y}(W)$: $\forall x(Xx \rightarrow \exists y(Yy \wedge Wxy))$
>
> $\text{Onto}_{X,Y}(W)$: $\forall y(Yy \rightarrow \exists x(Xx \wedge Wxy))$ (Donaldson MS: 7)

The effect of these definitions (and various background assumptions about grounding) is to proliferate grounding facts, due to the presence of the existential and universal quantifiers. For instance, Donaldson obtains the result that $2_{Card} \neq 0_{Card} \ll 1_{Card} = 1_{Card}$ Why? Because (on his assumptions about grounding and universal quantification) 2_{Card}'s not having the property $(\lambda x \,.\, x = 0_{Card})$ partially grounds the falsity of the antecedent, and thus the vacuous truth, of $\text{Total}_{X,Y}((\lambda xy \,.\, x = 0_{Card} \wedge y = 0_{Card})$, which in turn grounds $(\lambda x \,.\, x = 0_{Card}) \sim_{(\lambda xy.x=0_{Card} \wedge y=0_{Card})} (\lambda x \,.\, x = 0_{Card})$ by princi-

ples for conjunction, which in turn grounds $\exists W\,(\lambda x\,.\,x = \mathrm{o}_{Card}) \sim_W (\lambda x\,.\,x = \mathrm{o}_{Card})$ by principles for the existential quantifier.

Donaldson canvasses several responses to this problem, including accepting grounding regresses (MS: 12) and altering (SRP) by introducing restricted quantifiers (MS: 17). On my view, however, the problem is more basic: to maintain that equinumerosity must be cashed out at the fundamental level in terms of the standard sentential connectives and first- and second-order existential and universal quantifiers is to engage in connective chauvinism: the assumption that the small number of familiar connectives play a uniquely privileged role in metaphysical structure and that any more complex logical constants must ultimately be broken down to them. In ordinary cases, of course, connective chauvinism is harmless, because defined and primitive logical constants are necessarily extensionally equivalent. But necessary extensional equivalence is not good enough inside a grounding operator: grounding contexts are non-congruential, where necessarily extensionally equivalent expressions are not always substitutable *salva veritate*.[9]

Because HP is supposed to capture metaphysical structure, it is reasonable to assume that $\approx$ should be understood as picking out a logical higher-order relation by direct reference, not by means of a description containing several clauses.[10] Equinumerosity, after all, is logical on the standard Tarski–Sher–McGee notion of permutation-invariance discussed in §2.2. The extreme generality of the treatment of logical constants in $\mathscr{L}^{\Sigma}$ in that section was motivated by the desire to allow for such connectives to be implemented directly, rather than by definition (and we were careful to leave the question how to understand $\approx$, which is only relevant in contexts such as these, open when HP was introduced in Chapter III). But it is far from clear that a primitive

[9] On grounding and non-congruentiality, see Schnieder (2011: 453). The simplest example is self-conjunction: almost all grounding theorists accept $\phi \rightarrow (\phi \lessdot \phi \wedge \phi)$ and reject $\phi \rightarrow (\phi \lessdot \phi)$.

[10] An objection: have I not treated directly referential terms as intersubstitutable *salva veritate* in grounding contexts in my reply to Schwartzkopff? There is a crucial difference; there can be no distinct but coextensive type-$\overline{e}$ entities, so no failures of extensionality can arise from directly referential terms for individuals. But there can be distinct but coextensive type-$\overline{(e \rightarrow t), (e \rightarrow t) \rightarrow t}$ entities. That Muskens models appropriate model this (since $I_e = E_e$) is another reason to think that $\mathfrak{Mst}$ is weakly faithful to truth, as I argued in §2.3.

$\approx$ connective should behave like Donaldson's defined connective inside GHP; it might be thought, for instance, that equinumerosity is similar enough to identity that true propositions of the form $F = F$ are ungrounded.

This idea can certainly be resisted: even intuitions about grounding for the workaday connectives are highly contested; our insight into the grounding behaviour of a primitive equinumerosity constant can hardly be thought pellucid. This points, however, to another substantive issue. In §4.1, I did not use grounding or essence operators to derive claims about the dependence profiles of ordinals, but relied on intuitions about the latter claims directly. Ultimately, however imperfect our knowledge about objectual dependence and immediate objectual dependence, these are simply Relations among objects and can be treated in an ordinary extensional framework; in contrast, the vicissitudes of non-congruential operators engender complexity on an entirely different scale. Although I find the argument in §5.2 highly plausible, it is no succedaneum for the main dependence argument: to refuse to work with primitive objectual dependence and demand arguments in terms of grounding or essence instead is to impose a needless burden.

VI. HIGHER-ORDER ABSTRACTION PRINCIPLES

The main argument of Chapters I–V concluded with the claim that SCO is a well-motivated theory of sets, cardinals, ordinals, and natural numbers. But SCO uses only a fragment of the expressive power of the type theory developed in Chapter II: except for the abstraction operators, it never steps beyond $\mathscr{L}^2$, ordinary extensional second-order logic. It is natural to ask whether abstraction principles like the second-order principles used in SCO can be extended up the hierarchy. The only sustained discussion of third- and higher-order abstraction principles is due to Roy Cook (2009), who established that the E-tenability of certain higher order analogues of HP on various cardinals is independent of ZFC. In this chapter, I extend Cook's results, using the spectrum of semantic frameworks for type theory set out in §§2.2–4.

6.1 Basic Results: The Extensional Case

We first introduce a few abbreviations for convenience. We define the type $\langle 1^1 \rangle$ as e, $\langle 2^1 \rangle$ as $(e \to t)$, and we recursively define the type $\langle (i+1)^1 \rangle$ as $(\langle i \rangle \to t)$, and for $i \geq 2$, $\langle i^{n+1} \rangle := (\langle i-1 \rangle, \ldots, \langle i-1 \rangle \to t)$, with n occurrences in total of $\langle i-1 \rangle$. We omit the superscript n when $n = 1$.

In other words, type-$\langle i \rangle$ terms are singulary terms of ith order, and type-$\langle i^n \rangle$ terms ith-order n-ary terms. In this framework, the abstraction operators we are interested in will be of type $(\langle i^n \rangle \to \langle j^n \rangle)$ for $j < i$. We focus on two cases in particular: where $j = 1$, the abstraction operator is *deep*: it takes ith-order terms to first-order terms; where $j = i - 1$, the abstraction operator is *shallow*: it takes ith-order terms to $i-1$th-order terms. In the degenerate case when $i = 2$, the abstraction operator is both deep and shallow. These generate corresponding families of abstraction principles. A *deep higher-order abstraction principle (DHAP) family* is a schema of the form

$$\forall x^{\langle i \rangle} \forall y^{\langle i \rangle} (\S^{(\langle i \rangle \to e)} x = \S y \equiv \phi_i(x, y)), \tag{6.1.1}$$

for $i \geq 2$, where ϕ_{i+1} is generated from ϕ_i by uniform type-raising. A *shallow higher-order abstraction principle (SHAP) family* is a schema of the form

$$\forall x^{\langle i \rangle} \forall y^{\langle i \rangle} (\S^{(\langle i \rangle \to \langle i-1 \rangle)} x = \S y \equiv \phi_i(x, y)). \tag{6.1.2}$$

In both cases, we term the $i = 2$ instance the *base principle*.

Of course, many options exist for higher-order abstraction principles other than (6.1.1) and (6.1.2), but these are the most natural ways of extending principles such as HP or RITA to higher types. In particular, abstraction principles that map ith-order terms to jth-order terms for $1 < j < i-1$, whatever their intrinsic interest, appear highly *ad hoc* if conceived of as attempts to generalize a base principle.

In the case of HP, the DHAP family takes the form

$$\forall x^{\langle i \rangle} \forall y^{\langle i \rangle} (\#^{(\langle i \rangle \to e)} x = \#y \equiv x \approx_i y), \tag{D-HP^i}$$

where $\approx_i$ is the higher-order version of equinumerosity (i.e., the existence of a bijective ith-order function between ith-order entities). For present purposes, this can be taken to be either a primitive operator or a defined abbreviation for a higher-order formula using only standard logical constants.[1]

We can now state Cook's results and then proceed to novel ones:

Theorem 6.1.1 (Cook 2009: 359–60) Define $\beth_i(\kappa)$ as follows: $\beth_0(\kappa) = \kappa; \beth_{i+1}(\kappa) = 2^{\beth_i(\kappa)}$. D-HP^i is E-tenable on κ just in case $|\alpha| \leq \kappa$, where $\aleph_\alpha = \beth_i(\kappa)$.

It follows from this theorem that, if GCH holds, then each HP^i has a model of every cardinality. But it is also consistent with ZFC that, for all $i > 2$, no model exists of D-HP^i (Cook 2009: 361). Thus, Cook obtains the following independence result:

[1] Specifically, define ith-order injection as $\mathrm{Inj}_i(f^{(\langle i-1 \rangle \to \langle i-1 \rangle)}, x^{\langle i \rangle}, y^{\langle i \rangle}) \stackrel{\text{Df.}}{=} \forall z^{\langle i-1 \rangle}(xz \to yf(z))$, ith-order bijection as $\mathrm{Bij}_i(f^{(\langle i-1 \rangle \to \langle i-1 \rangle)}, x^{\langle i \rangle}, y^{\langle i \rangle}) \stackrel{\text{Df.}}{=} \mathrm{Inj}_i(f^{(\langle i-1 \rangle \to \langle i-1 \rangle)}, x^{\langle i \rangle}, y^{\langle i \rangle}) \wedge \mathrm{Inj}_i(f^{(\langle i-1 \rangle \to \langle i-1 \rangle)}, y^{\langle i \rangle}, x^{\langle i \rangle})$, and ith-order equinumerosity as $x^{\langle i \rangle} \approx_i y^{\langle i \rangle} \stackrel{\text{Df.}}{=} \exists f^{(\langle i-1 \rangle \to \langle i-1 \rangle)} \mathrm{Bij}(f, x, y)$. (Some type superscripts have been elided for readability.)

Theorem 6.1.2 (Cook 2009: 361) The E-satisfiability of D-HP^i for $i > 2$ is independent from ZFC.

We now consider the DHAP families corresponding to NEW V, RITA, and SOAP.

The DHAP family corrresponding to NEW V takes the form:

(D-NEW V^i) $\quad \forall x^{\langle i\rangle}\forall y^{\langle i\rangle}(\in^{(\langle i\rangle\to e)}x = \in y \equiv (\mathrm{Big}_i^{\langle i+1\rangle}x \wedge \mathrm{Big}_i y) \vee \forall z^{\langle i-1\rangle}(xz \equiv yz))$.

We can define the ith-order analogue of bigness, for $i \geq 2$, as follows:

(6.1.3) $$\mathrm{Big}_i^{\langle i+1\rangle}x^{\langle i\rangle} \stackrel{\text{Df.}}{=} x \approx_i (\lambda z^{\langle i-1\rangle}\,.\, z = z).$$

As noted earlier, NEW V and HP are proof-theoretically closely related: PA^2 and FN (the theory whose only axiom is New V) are interpretable in each other (Burgess 2005: 189). And if GCH holds, then NEW V is satisfiable on every regular cardinal (Shapiro and Weir 1999: 317–18). So it might be suspected that the model theory of D-NEW V^i should be similar to that of D-HP^i. This is, however, not the case.

Theorem 6.1.3 D-NEW V^i is E-unsatisfiable for $i \geq 3$.

Proof. We show the $i = 3$ case; for $i > 3$, a precisely parallel argument applies. Assume for *reductio* that D-NEW V^3 is satisfiable with a model where $|D_e| = \kappa$. Let Q denote the class of equivalence classes (under coextensiveness) of non-Big$_3$ entities in $D_{\langle 3\rangle}$ of this model. Thus there is an injection from Q to D_e. Let $Z = \{x \in D_{\langle 3\rangle} : \exists y \in D_{\langle 2\rangle} x = \{y\}\}$. Each $z \in Z$ is smaller than universe-sized and each has a different extension; thus no two will lie in the same equivalence class in Q, so $|Q| \geq |Z|$. But $|Z| = |D_{\langle 2\rangle}| = 2^\kappa$. So $|Q| \geq 2^\kappa$ and, by Cantor's theorem, there is no injection from Q to D_e. ⊣

The DHAP family corrresponding to RITA^n takes the form

$$(\text{D-RITA}^{n(i)}) \quad \forall x^{\langle i^n\rangle}\forall y^{\langle i^n\rangle}(\Omega^{(\langle i^n\rangle\to e)}x = \Omega y \equiv (\text{Big}_i^{\langle i+1\rangle}\text{Fld}_{i,n}^{(\langle i^n\rangle\to i)}x \wedge (\text{Big}_i\text{Fld}_{i,n}y) \vee \wedge x \cong_{i,n}^{(\langle i^n\rangle,\langle i^n\rangle\to t)} y))).$$

To show that this too is unsatisfiable, we prove a stronger result. The DHAP family that corresponds to SOAP is

$$(\text{D-SOAP}^i) \quad \forall x^{\langle i^2\rangle}\forall y^{\langle i^2\rangle}(\Omega^{(\langle i^2\rangle\to e)}x = \Omega y \equiv (\neg\text{WO}_i^{(\langle i^2\rangle\to t)}x \vee \text{Big}_i^{\langle i+1\rangle}\text{Fld}_i^{(\langle i^2\rangle\to i)}x) \vee (\neg\text{WO}_i y \vee \text{Big}_i\text{Fld}_i y) \vee ((\neg\text{Big}_i x \wedge \neg\text{Big}_i y \wedge \text{WO}_i x \wedge \text{WO}_i y) \wedge x \cong_i^{(\langle i^2\rangle,\langle i^2\rangle\to t)} y)),$$

where WO_i, Fld_i, and $\cong_i$ represent the natural ith-order generalizations of WO, Fld, and $\cong$. SOAP is as well-behaved, model-theoretically, as HP: it has a model of every infinite cardinality. Nonetheless, its DHAP analogues are also unsatisfiable.

THEOREM 6.1.4 D-SOAP^i is E-unsatisfiable for $i \geq 3$.

Proof. Once again, we consider the $i = 3$ case; extension to $i \geq 3$ is trivial. Assume for *reductio* that D-SOAP3 is satisfiable with a model where $|D_e| = \kappa$. Let Q denote the class of equivalence classes (under isomorphism) of relations in $D_{\langle 3\rangle^2}$ of this model that have non-Big_3 fields and well-order their fields. Thus there is an injection from Q to D_e. Define $\zeta : D_{\langle 2\rangle^2} \to D_{\langle 3\rangle^2}$ so that $\langle\{x\},\{y\}\rangle \in \zeta(a)$ if and only if $\langle x, y\rangle \in a$. Clearly, for $a, b \in D_{\langle 2\rangle^2}$, a is isomorphic to b if and only if $\zeta(a)$ is isomorphic to $\zeta(b)$. Define $Z = \{\zeta(a) : a \in D_{\langle 2\rangle}$ and a is a well-ordering$\}$. Every member of Z has a field of size at most as large as D_e; thus, none is Big_3. But there are κ^+ nonisomorphic well-orderings over a set of size κ, so there are κ^+ nonisomorphic relations in Z with non-Big_3 fields, no two of which will be mapped to the same equivalence class in Q. So $\|Q\| \geq \kappa^+$. Thus there is no injection from Q to D_e. ⊣

COROLLARY 6.1.5 D-RITA$^{n(i)}$ is E-unsatisfiable for all $n \geq 2$, $i \geq 3$.

Proof. An E-model $\mathfrak{M}$ of D-SOAP^i over D_i could be generated from a model $\mathfrak{N}$ of D-RITA^i over D_i simply by defining $v_c^{\mathfrak{M}}(\Omega) : D_{\langle i^2 \rangle} \to D_e$ as follows: $v_c^{\mathfrak{M}}(\Omega) : x \mapsto v_c^{\mathfrak{N}}(\Omega)(x)$ when x is (the higher-order generalization of) a well-ordering, and $v_c^{\mathfrak{M}}(\Omega) : x \mapsto v_c^{\mathfrak{N}}(\Omega)(D_{\langle i-1 \rangle} \times D_{\langle i-1 \rangle})$ otherwise. It follows from this that D-RITA$^{i(n)}$ is satisfiable for no $n \geq 2$ for any $i \geq 3$, since a model of such a D-RITA$^{i(n)}$ could similarly be turned into a model of D-RITA^i simply by collapsing excess argument places. ⊣

6.2 Philosophical Implications of the Extensional Case

How problematic are Theorems 6.1.1–5 for someone who wishes to defend the *base* (second-order) versions of RITA^n (or, for that matter, HP, SOAP, or NEW V)? There are two philosophical arguments, with different dialectical stakes, that can be mounted using the technical results about DHAP families against the corresponding base principles.

The first argument, discussed by Cook (2009: 363–68), is specific to Theorem 6.1.2. It relies on a distinctly *neologicist* conception of the role of abstraction principles—of the type I argued against in §3.2—on which the only way we can come to understand the semantic value of the # operator is by grasping HP. The argument is a version of the bad company objection, familiar from Boolos (1997). If an abstraction principle is to do any useful mathematical work, then it cannot meet the standard conservativeness and eliminability requirements on an admissible definition. The neologicist needs to give some principled account of the conditions under which an abstraction principle is acceptable: the acceptability criterion must be weaker than conservativeness and eliminability, but strong enough to avoid allowing clearly unkosher abstraction principles. The bad company objection arises because obviously untenable abstraction principles, such as Basic Law V, are superficially quite similar to HP—by the neologicist's lights, the paradigm case of a good abstraction principle. Thus, whatever the right acceptability criterion is, it will not be something that can simply be read off the surface form of a principle. Moreover, because abstraction principles are supposed to provide a way of *understanding* the abstraction operator, the acceptability criterion will have to be in some sense cognitively *tractable*. Tractability need not imply the existence of

an effective decision procedure: a bare consistency criterion is arguably tractable, even though not effectively decidable. Bare consistency, however, is ruled out as a criterion both by the existence of individually consistent but pairwise inconsistent abstraction principles (Wright 1997) and by consistent abstraction principles that are unacceptable for other reasons (such as those which constrain the size of the universe in a way incompatible with standard mathematics).

For the neologicist, as Cook argues, Theorem 6.1.2 offers "a new version of the Bad Company objection, one with a new twist not present in the earlier sequence of worries and responses" (2007: 363). He formulates the argument using irenicity[2]—compatibility with any conservative abstraction principle, a much-discussed candidate acceptability criterion—but the problem is more general; we can put it in terms of the following inconsistent tetrad, with which the neologicist is confronted:

(6.2.1) If neologicism is correct, then for every plausible candidate abstraction principle $\mathscr{A}$, there exists a tractable criterion for $\mathscr{A}$'s acceptability.

(6.2.2) HP is a plausible candidate abstraction principle.

(6.2.3) If HP is a plausible candidate abstraction principle, then so is D-HP^i for $i > 2$.

(6.2.4) There is no tractable criterion for the acceptability of HP^i for $i > 2$.

None of these is easy for the neologicist to give up. Whatever the precise bounds of tractability are, if the question of HP^i's acceptability for $i > 2$ turns solely on satisfiability, and this turns on questions of set theory unsettled by ZFC and intimately related to the Continuum Hypothesis, then the question is surely intractable. The only way to maintain tractability will be if there exists some independent ground other than unsatisfiability for rejecting higher-order analogues of HP. But it is not obvious what this ground could be, unless it amounts simply to the rejection of all higher-order analogues

[2]On irenicity and other standard candidate criteria for abstraction principles, see Cook (2009: 352–355, 2010) and Eklund (2009).

of abstraction principles. And the considerations that motivate (6.2.3) tell against that constraint: as Cook notes, it is very hard to see why, if counting objects is coherent, counting concepts should not be. The novel results given in Theorems 6.1.3–5, however, do not pose a similar problem for the neologicist: an explicit *proof* of unsatisfiability provides grounds for declaring that $\mathscr{A}$ is unacceptable that are tractable by any standard.

Cook's argument, however, does not threaten the version of AAR I have defended in this work. On my view, whether we should incorporate an abstraction principle into our account of the foundations of mathematics depends on what it can contribute to our best general total theory of the mathematical realm, and these total theories are to be evaluated using ordinary canons of theory choice—explanatory power, simplicity, elegance, and so on. But there is no guarantee that these canons will yield success: just as there may be laws of physics which we are in no position to know, there may be true abstraction principles which we are unable to discover. (And likewise for basic principles in set theory, mereology, the theory of metaphysical modality, and so on: perhaps GCH or its negation is itself such a principle.) So there is no reason to expect tractable acceptability conditions for abstraction principles, and the tetrad of (6.2.1–4) is not a threat.

The proponent of AAR, however, does face a problem due to Theorems 6.1.3–5, because of the importance of generalizability in the principles of theory choice of which she avails himself. *Ceteris paribus*, a second-order principle which has a natural extension to higher types is more elegant and powerful than one which does not; thus Theorem 6.1.5, it seems, must in itself count against the overall theoretical merits of any system built around the base case of RITA^n (and *mutatis mutandis* for SOAP and NEW V). (By the same reasoning, Theorem 6.1.2 will also turn out to be evidence against HP if in fact we should discover that, in virtue of a gross violation of GCH, there is some i such that D-HP^i is unsatisfiable.)

But there are two reasons that these results should not cause the proponent of abductive methodology *greatly* to lower her credence in a theory built around RITA^n, de-

fended here. First, no competing non-abstractionist theory of cardinals and ordinals avoids these limitations: analogous theorems will apply, for instance, to attempts to extend set-theoretic definitions of cardinals and ordinals to higher types. Second, the SHAP families provide an important respect in which, given appropriate additional assumptions, RITA^n (and SOAP and NEW V) *are* generalizable to higher types, subject to a caveat to be discussed in the next section. Let s-HP^i, s-NEW V^i, s-SOAP^i, and s-RIT$A^{(n)i}$ be the families generalizing NEW V, SOAP, and HP in accordance with schema (6.1.2); they differ from the DHAP families only by replacing an operator $\S^{(i\to e)}$ or $\S^{(i^n\to e)}$ with $\S^{(i\to i-1)}$ or $\S^{(i^n\to i-1)}$ (and, implicitly, retyping the identity sign on the left-hand side):

(s-HP^i) $$\forall x^{\langle i\rangle}\forall y^{\langle i\rangle}(\#^{(\langle i\rangle\to\langle i-1\rangle)}x=\#y\equiv x\approx_i^{\langle(i+1)^2\rangle}y),$$

(s-NEW V^i) $$\forall x^{\langle i\rangle}\forall y^{\langle i\rangle}(\in^{(\langle i\rangle\to\langle i-1\rangle)}x=\in y\equiv ((\mathrm{Big}_i^{\langle i+1\rangle}x\wedge\mathrm{Big}_i y)\vee\forall z^{\langle i-1\rangle}(xz\equiv yz)),$$

(s-SOAP^i) $$\forall x^{\langle i^2\rangle}\forall y^{\langle i^2\rangle}(\Omega^{(\langle i^2\rangle\to\langle i-1\rangle)}x=\Omega y\equiv((\neg\mathrm{WO}_i^{(\langle i^2\rangle\to t)}x\vee\mathrm{Big}_i^{\langle i+1\rangle}\mathrm{Fld}_i^{(\langle i^2\rangle\to i)}x)\vee(\neg\mathrm{WO}_i y\vee\mathrm{Big}_i\mathrm{Fld}_i y)\vee((\neg\mathrm{Big}_i x\wedge\neg\mathrm{Big}_i y\wedge\mathrm{WO}_i x\wedge\mathrm{WO}_i y)\wedge x\cong_{i,2}^{(\langle i^2\rangle,\langle i^2\rangle\to t)}y))),$$

(s-RIT$A^{n(i)}$) $$\forall x^{\langle i^n\rangle}\forall y^{\langle i^n\rangle}(\Omega^{(\langle i^n\rangle\to\langle i-1\rangle)}x=\Omega y\equiv(\mathrm{Big}_i^{\langle i+1\rangle}\mathrm{Fld}_{i,n}^{(\langle i^n\rangle\to i)}x\wedge(\mathrm{Big}_i\mathrm{Fld}_{i,n}y)\vee\wedge x\cong_{i,n}^{(\langle i^n\rangle,\langle i^n\rangle\to t)}y))).$$

THEOREM 6.2.1 s-HP^i (resp. s-NEW V^i, s-SOAP^i, s-RIT$A^{n(i)}$) is E-tenable on κ if and only if s-HP^i (resp. s-NEW V^i, s-SOAP^i, s-RIT$A^{n(i)}$) is tenable on λ, where $\kappa=\beth_{i-2}(\lambda)$.

Proof. By a simple model construction, establishing a bijection between $D_{\langle i\rangle}$ of a model $\mathfrak{M}$ for s-HP^i or s-NEW V^i ($D_{\langle i^2\rangle}$ or $D_{\langle i^n\rangle}$ in the case of s-SOAP^i or s-RIT$A^{n(i)}$), where

$|D_e(\mathfrak{M})| = \kappa$, and $D_{\langle 2 \rangle}$ of a model $\mathfrak{M}'$ for NEW V or HP ($D_{\langle 2^2 \rangle}$ in the case of SOAP, $D_{\langle 2^n \rangle}$ in the case of RITA^n), where $|D_e(\mathfrak{M}')| = \lambda$. ⊣

COROLLARY 6.2.2 If GCH holds, then for all i and all n, S-RITA$^{n(i)}$ is E-tenable on every κ.

The picture of generalization to which the SHAP principles correspond is different from that of the DHAP principles. The SHAP principles can be said to capture the intuition of *typical ambiguity*: the idea that reinterpretation of all type symbols up or down the hierarchy should not affect truth value. Such wholesale reinterpretation would affect constants, including generalized abstraction operators, as much as variables. The DHAP principles, in contrast, reflect a *nominalizability* intuition: the idea that a uniform principle should be available allowing us to find objects corresponding to higher-order entities of any type. But of these two, the first is the more reasonable desideratum. Cantor's theorem means that we shall *never* be able to have a consistent theory of nominalization as free-wheeling as natural language, where we imagine ourselves free to iterate 'the property that a ... possesses if ...' and kindred constructions at will, would lead us to expect. And Cantor's theorem also means that the deficit grows worse as the hierarchy of types is ascended: at each stage, *more and more* higher-type entities have to be shoehorned into appropriate equivalence classes for a type-lowering function to be possible.

In fine, the lesson we should draw from these results is the following: because type-dropping principles face cardinality constraints that increase with the depth of the drop, we should not normally expect DHAP-style generalization to be a feasible theoretical goal. The unsatisfiability of D-RITA$^{n(i)}$ and kindred principles proceeds from the same cause as the inconsistency of Basic Law V itself.

6.3 M-Models and Nonextensionality

This argument does not settle the question, however, whether we ought to accept a DHAP when we can get it: for instance, should we expect that, if GCH holds, D-HP^i

is true on its intended interpretation? This is a difficult philosophical problem, but there is a technical result that suggests caution: *if higher-order extensionality fails badly enough,* D-HP^i *may not be tenable on the universe even given GCH.*

The central issue here is that $\mathfrak{SEL}$ is not, in the terminology of §2.2, weakly faithful to truth: there are principles (such as $\forall X \forall Y(X \equiv Y \leftrightarrow \forall x(Xx \leftrightarrow Yx))$, to choose the simplest example) that hold in every SE-model but fail on the intended interpretation. In contrast, I argued that $\mathfrak{Nfl}$ is weakly faithful to truth. If a given abstraction principle were strongly M-tenable on every κ with a certain size property (say, inaccessibility or measurability), this would be no guarantee of its truth on the intended interpretation, but it would mean that there is no *purely formal* constraint other than the number of objects. (Of course, there is no guarantee that we could come to know this: for instance, it might, like the E-tenability of D-HP^i, turn on difficult questions of cardinal exponentiation.)

Unfortunately, we are not so lucky:

Theorem 6.3.1 For $i > 2$, D-HP^i is strongly M-tenable on no κ.

Proof. We prove the result for D-HP^3. As before, extension to other cases is simple. If κ is finite, the claim is trivial; thus we construct an M-frame $\mathscr{F}$ of arbitrary infinite cardinality that cannot be extended so as to satisfy D-HP^3. Choose disjoint sets A and B, neither of which has 0 or 1 as an element, such that $|A| = \kappa$ and $|B| = \lambda$, where $\lambda \geq 2^\kappa$ and $\lambda \geq \aleph_{\kappa^+}$. (Elementary set theory ensures that such sets will exist, whether or not GCH holds.) Now define the M-protoframe $I_{\mathscr{T}}$ as follows: let $I_e = A$; $I_t = \{0, 1\}$; $I_{(e \to t)} = B$; for all $\tau \in \mathscr{T}$ other than these, $I_\tau = I_{(\tau_1, \ldots, \tau_n \to \tau_0)} = {}^{I_{\tau_0}}(I_{\tau_i} \times \cdots \times I_{\tau_n})$. Choose an arbitrary η such that $\mathscr{F} = \langle I_{\mathscr{T}}, \eta \rangle$ is full. (Since $\lambda \geq 2^\kappa$, such an η must exist.) For $\kappa = \aleph_\lambda$, let $\partial(\kappa) = \aleph_0 + \lambda$. (In other words, $\partial(\kappa)$ is the number of equivalence classes under equinumerosity of a κ-membered collection). An interpretation function v such that $\mathfrak{M} = \langle \mathscr{F}, v \rangle$ and $\mathfrak{M} \vDash$ D-HP^3 can be found only if there exists an injective $f : \partial(|B|) \to A$. But $|B| \geq \aleph_{\kappa^+}$, so $\partial(|B|) \geq \kappa^+$. ⊣

The result exploits the fact that, with only weak constraints on the relative cardinality

of the domains in an M-protoframe, $|I_\tau|$ can massively exceed $|E_\tau|$. The result also depends on the fact that the equivalence relation in D-HP^3 operates *directly* on entities of higher type than the target type of the abstraction functor. No such result is possible for S-RITA$^{n(i)}$, for instance, since the abstraction functor maps any two type-$\langle i^n \rangle$ entities that are isomorphic into the same equivalence class, and whether or not isomorphism between (Super-Super-...)-Superrelations holds is wholly a function of the *extension* of those (Super-Super-...)-Superrelations.

Of course, we might obtain additional evidence that these M-models are not worth considering. This suggests a principle which, if true (and cardinal arithmetic cooperates), will ensure that the universe does not have the pathological property of the model constructed in Theorem 6.3.1; informally, it will ensure that the higher-order structure of reality 'corresponds' to a model that *can* be extended appropriately, assuming other conditions are in place. First we define a model class:

Definition 6.3.2 An M-model is *Concept-noninflationary* just in case, for all $\tau \in \mathscr{T}$ such that $\tau \neq t$, $|I_\tau| = |E_\tau|$. An M-model is *proposition-noninflationary* just in case $|I_t|^{|I_e|} \leq \beth_2(|I_e|)$. An M-model is noninflationary if it is both Concept-noninflationary and proposition-noninflationary. We use $\mathfrak{Noninf}$ to denote the class of noninflationary M-models.

If $\partial(\beth_2(\kappa)) < \kappa^+$, then D-HP^i will be strongly tenable over $\mathfrak{Noninf}$ for κ just in case it is weakly tenable over $\mathfrak{Noninf}$ for κ.

The thesis that interests us is:

(NONINFLATION) $\mathfrak{Noninf}$ is weakly faithful to truth.

Is NONINFLATION true? This raises a number of difficult questions, which exceed the scope of this project.

Here is, however, one fairly straightforward consideration in favour of a restriction to Concept-noninflationary M-models. On the *aboutness* picture, developed by Jeremy Goodman (MS), a Concept is uniquely determined by the combination of a *qualitative profile* and a *plurality*, possibly proper-class-sized, of objects that the proposition

is *about*. We focus on the $(e \to t)$ case for simplicity. The intuitive idea is that, for example, a purely qualitative Concept F, the Concept $(\lambda x \,.\, Fx \wedge a = a)$, the Concept $(\lambda x \,.\, Fx \wedge b = b)$, and the Concept $(\lambda x \,.\, Fx \wedge a = a \wedge b = b)$ are pairwise distinct, although coextensive. (Coextensiveness here simply means taking the same objects to the same propositions; since we are allowing that $\overline{\mathscr{D}_t}$ contains more than two propositions, even in the absence of distinct but coextensive Concepts, the framework need not be 'extensional' in the ordinary sense.)

The first is about no particular object (since it is purely qualitative), the second about a, the third about b, and the fourth about both a and b—and so on, *ad transfinitum*. If the theory of aboutness is true and qualitative profiles are no more numerous than functions from objects to propositions, then there will be no more distinct coextensive Concepts than there are pluralities of objects. In the framework of M-models, this means that, if $|I_e| = \kappa$ and $|I_t| = \lambda$, then $|E_{(e\to t)}| = \lambda^\kappa$, but $|I_{(e\to t)}| = 2^\kappa \cdot \lambda^\kappa = \lambda^\kappa$ (since $\lambda \geq 2$).

This does not, however, establish NONINFLATION, for it must still be the case that $\lambda^\kappa \leq \beth_2(\kappa)$; if not, then an M-frame might be incapable of extension to an M-model satisfying D-HP^3 not because of the existence of a multiplicity of coextensive but distinct Concepts but simply because the class of propositions is large enough to swamp the class of objects—in other words, $\partial(\lambda^\kappa) > \aleph_\kappa$. (The requirement that $\lambda^\kappa \leq \beth_2(\kappa)$ is chosen to allow for theories on which propositions are sets of worlds, provided that there are fewer worlds than objects. $\lambda^\kappa \leq \beth_2(\kappa)$ may be far stronger than the requirement that $\lambda^\kappa \leq \partial(\kappa)$, depending on the facts of cardinal exponentiation, but it is hard to find a principled reason to tolerate just enough propositional promiscuity to drive D-HP^3 to the edge of unsatisfiability.)

Is this second requirement for $\mathfrak{Noninf}$'s faithfulness met? This turns on very deep general questions of propositional fineness of grain—an area in which philosophical research, despite Cian Dorr's sweeping contribution (2016), is still in an early stage. The idea that $|I_t|^{|I_e|} = 2^{|I_e|}$ is appealing, for it imposes order on some of the sprawl of $\mathfrak{Msk}$, but as yet I have found no convincing argument for it. The question, and related

questions about faithful model classes in a non-extensional framework, remains open.

CONCLUSION

David Lewis famously observed:

> Philosophical theories are never refuted conclusively. (Or hardly ever: Gödel and Gettier may have done it.) The theory survives its refutation—at a price. Argle has said what we accomplish in philosophical argument: we measure the price. (Lewis 1983–86: 1.x[1])

The Kinds of Mathematical Objects is no exception to this. I do not claim to have made a decisive case for antireductionism about cardinals and ordinals. I hope to have shown, however, that the price of reductionism is higher and that of antireductionism lower than one might antecedently have thought. I began by arguing that antireductionism does not suffer from what many philosophers take (correctly) to be an exorbitantly expensive flaw—incompatibility with mathematical practice. Instead, mathematics is a complex and multifarious enterprise that does not tell strongly for or against reductionism. But, I argued, the cost of reductionism does not lie where many philosophers have located it—in Benacerraf's multiple realizability argument. For metasemantic reasons, this argument is not particularly persuasive. Its actual cost lies in its conflict with very plausible dependence theses.

If the dependence argument stood on its own, the theoretical balance might well still tip towards reductionism: we should not abandon a reductionist account that is, in strictly mathematical terms, adequate on the basis of contested metaphysical intuitions unless we have something equally good on hand to replace it. But the positive proposal motivated by abductivist abstractionist realism, including the theory SCO, is at least as elegant and powerful as standard set theory, and it recapture every mathematically relevant result. These considerations, taken in sum, probably do not amount to enough for us to be able to *know* that antireductionism is true. But they should, at least, cause us to think it more likely than not to be true.

[1] Footnotes omitted.

In addition to arguing for antireductionism, this work has a secondary purpose: demonstrating the utility of non-extensional type theory (and in particular the system of Muskens frames and models) for metaphysical inquiry. Here much remains to be done: the tools are powerful enough, but we do not yet know how to apply them to philosophical problems about the structure of higher-type reality (particularly at third order and above, where hyperintensionality becomes crucial), as the difficulty involved in assessing NONINFLATION demonstrates. If nothing else, however, the potential of the formal apparatus is clear. "There is no mathematical substitute for philosophy" (Kripke 1976: 416), including philosophy of mathematics, but there is also no non-mathematical substitute for the mathematical tools philosophy requires.

BIBLIOGRAPHY

When originals and reprints are listed in the same entry, an asterisk indicates the edition used for pagination in citations.

Ackerman, Diana (1978). '*De Re* Propositional Attitudes Toward Integers'. *Southwestern Journal of Philosophy* **9.2**: 145–53.

Aczel, Peter (1988). *Non-Well-Founded Sets.* Stanford: Center for the Study of Language and Information.

Almog, Joseph, John Perry and Howard Wettstein (1989). *Themes from Kaplan.* Oxford: Oxford University Press.

Anderson, C. Anthony (1977). 'Some Models for the Logic of Sense and Denotation with an Application to Alternative (0)'. Ph.D. dissertation, University of California, Los Angeles.

———— (1980). 'Some New Axioms for the Logic of Sense and Denotation: Alternative (0)'. *Noûs* **14**: 217–34.

———— (1984). 'General Intensional Logic'. In Gabbay and Guenther (1983–87): **2**.355–85.

———— (1987). 'Semantical Antinomies in the Logic of Sense and Denotation'. *Notre Dame Journal of Formal Logic* **28**: 99–114.

———— (1998). 'Alonzo Church's Contributions to Philosophy and Intensional Logic'. *Bulletin of Symbolic Logic* **4**: 129–71.

Andjelković, Miroslava, and Timothy Williamson (2000). 'Truth, Falsity, and Borderline Cases'. *Philosophical Topics* **28**: 211–44.

Antos, Carolin, Sy-David Friedman, Radek Honzik, and Claudio Ternullo (2015). 'Multiverse Conceptions in Set Theory'. *Synthese* **192**: 2463–88.

Armour-Garb, Bradley, ed. (2017). *Reflections on the Liar.* Oxford: Oxford University Press.

Armstrong, D.M. (1978). *Universals and Scientific Realism.* 2 vols. Cambridge: Cambridge University Press.

Arrigoni, Tatiana, and Sy-David Friedman (2013). 'The Hyperuniverse Program'. *Bulletin of Symbolic Logic* **19**: 77–96.

Audi, Paul (2012a). 'A Clarification and Defense of the Notion of Grounding'. In Correia and Schnieder (2012b): 101–21.

——————— (2012b). 'Grounding: Toward a Theory of the In-Virtue-Of Relation'. *Journal of Philosophy* **109**: 685–711.

Auerbach, David (1994). 'Saying It with Numerals'. *Notre Dame Journal of Formal Logic* **35**: 130–46.

Awodey, Steve (1996). 'Structure in Mathematics and Logic: A Categorical Perspective'. *Philosophia Mathematica* (3rd ser.) **4**: 209–37.

Baker, Lynne Rudder (2000). *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.

Barwise, Jon (1975). *Admissible Sets and Structures: An Approach to Definability Theory*. Berlin: Springer.

———————, ed. (1974). *Handbook of Mathematical Logic*. Amsterdam: North-Holland.

———————, and Lawrence Moss (1996). *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena*. Stanford: Center for the Study of Language and Information.

Batchelor, Roderick (2010). 'Grounds and Consequences'. *Grazer Philosophische Studien* **80**: 65–77.

Beall, Jc, ed. (2003). *Liars and Heaps: New Essays on Paradox*. Oxford: Oxford University Press.

Benacerraf, Paul (1965). 'What Numbers Could Not Be'. *Philosophical Review* **74**: 47–73.

——————— (1973). 'Mathematical Truth'. *Journal of Philosophy* **70**: 661–79.

——————— (1985). 'Skolem and the Skeptic'. *Proceedings of the Aristotelian Society Supplement* **59**: 85–115.

——————— (1996). 'Recantation, or Any Old ω-Sequence Would Do After All'. *Philosophia Mathematica* (3rd ser.) **4**: 184–89.

———————— (1998). 'What Mathematical Truth Could Not Be I'. In Schirn (1998): 33–75.

————————, and Hilary Putnam (1964). *Philosophy of Mathematics*. Englewood Cliffs, N.J.: Prentice-Hall.

Bennett, Karen (2013). 'Having a Part Twice Over'. *Australasian Journal of Philosophy* **91**: 83–103.

Benzmüller, Christoph, Chad E. Brown, and Michael Kohlhase (2004). 'Higher-Order Semantics and Extensionality'. *Journal of Symbolic Logic* **69**: 1027–88.

Berger, Alan, ed. (2011). *Saul Kripke*. Cambridge: Cambridge University Press.

Bird, Alexander (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Birkhoff, Garrett, and Saunders Mac Lane (1977). *A Survey of Modern Algebra*. 4th edn. New York: Macmillan.

Boghossian, Paul, and Christopher Peacocke, eds. (2000). *New Essays on the A Priori*. Oxford: Clarendon Press.

Boolos, George (1971). 'The Iterative Conception of Set'. *Journal of Philosophy* **68**: 215–32. Rpt. in Boolos (1998): 13–29.*

———————— (1984). 'To Be is to Be a Value of a Variable (or to be Some Values of Some Variables)'. *Journal of Philosophy* **81**: 430–49. Rpt. in Boolos (1998): 54–72.*

———————— (1985). 'Nominalist Platonism'. *Philosophical Review* **94**: 327–44. Rpt. in Boolos (1998): 73–87.*

———————— (1989). 'Iteration Again'. *Philosophical Topics* **17**: 5–21. Rpt. in Boolos (1998): 88–104.*

———————— (1997). 'In Hume's Principle Analytic?' In Heck [1997]: 245–62. Rpt. in Boolos (1998): 301–14.

———————— (1998). *Logic, Logic, and Logic*. Intro. by John P. Burgess. Ed. Richard Jeffrey. Cambridge, Mass.: Harvard University Press.

Bricker, Phillip (1989). 'Quantified Modal Logic and the Plural *De Re*'. *Midwest Studies in Philosophy* **14**: 372–94.

Bueno, Otávio, and Øystein Linnebo, eds. (2009). *New Waves in the Philosophy of Mathematics.* Houndmills, Hants: Palgrave Macmillan.

Burgess, John (1983). 'Why I Am Not a Nominalist'. *Notre Dame Journal of Formal Logic* **24**: 93–105. Rpt. in Burgess (2008): 31–45.*

——————— (1998). 'Quinus ab omne naevo vindicatus'. *Canadian Journal of Philosophy Supplement* **23**: 25–66. Rpt. in Burgess (2008): 203–35.*

——————— (2005). *Fixing Frege.* Princeton: Princeton University Press.

——————— (2008). *Mathematics, Models, and Modality: Selected Philosophical Essays.* Cambridge: Cambridge University Press.

——————— (2015). *Rigor and Structure.* Oxford: Oxford University Press.

———————, A.P. Hazen, and David Lewis (1991). 'Appendix on Pairing'. In Lewis (1991): 121–49.

———————, and Gideon Rosen (1997). *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics.* Oxford: Clarendon Press.

Button, Tim (2010). 'Dadaism: Restrictivism as Militant Quietism'. *Proceedings of the Aristotelian Society* **110**: 387–98.

Butts, Robert E., and Jaakko Hintikka (1977). *Logic, Foundations of Mathematics, and Computability Theory: Proceedings of the Fifth International Congress of Logic, Methodology, and Philosophy of Science, Part I.* Dordrecht: Reidel.

Cantor, Georg (1878). 'Ein Beiträg zur Mannigfaltigkeitslehre'. *Crelles Journal für Mathematik* **84**: 242–58. Rpt. in Cantor (1932): 119–33.*

——————— (1883). *Grundlagen einer Allgemeinen Mannigfaltigkeitslehre.* Leipzig: B.G. Teubner. Rpt. in Cantor (1932): 165–209.*

——————— (1895). 'Beiträge zur Begründung der transfiniten Mengenlehre I'. *Mathematische Annalen* **46**: 481–512. Rpt. in Cantor (1932): 282–311. Trans. in Cantor (1915): 85–136.*

——————— (1897). 'Beiträge zur Begründung der transfiniten Mengenlehre II'. *Mathematische Annalen* **49**: 207–46. Rpt. in Cantor (1932): 312–56. Trans. in Cantor (1915): 137–201.*

———————— (1915). *Two Contributions to the Founding of the Theory of Transfinite Numbers.* Trans. P.E.B. Jourdain. Chicago: Open Court.

———————— (1932). *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts.* Ed. Ernst Zermelo. Berlin: Springer.

Cartwright, Richard (1994). 'Speaking of Everything'. *Noûs* **28**: 1–20.

Chalmers, David (2006). 'The Foundations of Two-Dimensional Semantics'. In García-Carpintero and Macià (2006): 55–140.

Cheyne, Colin (2001). *Knowledge, Cause, and Abstract Objects: Causal Objections to Platonism.* Dordrecht: Kluwer.

Church, Alonzo (1940). 'A Formulation of the Simple Theory of Types'. *Journal of Symbolic Logic* **5**: 56–68.

———————— (1946). 'A Formulation of the Logic of Sense and Denotation (Abstract)'. *Journal of Symbolic Logic* **11**: 31.

———————— (1951a). 'A Formulation of the Logic of Sense and Denotation'. In Henle, Kallen, and Langer (1951): 3–24.

———————— (1951b). 'The Need for Abstract Entities in Semantic Analysis'. *Proceedings of the American Academy of Arts and Sciences* **80**: 100–12.

———————— (1954). 'Intensional Isomorphism and Identity of Belief'. *Philosophical Studies* **5**: 65–73.

———————— (1973). 'Outline of a Revised Formulation of the Logic of Sense and Denotation (Part I)'. *Noûs* **7**: 24–33.

———————— (1974). 'Outline of a Revised Formulation of the Logic of Sense and Denotation (Part II)'. *Noûs* **8**: 135–56.

———————— (1993). 'A Revised Formulation of the Logic of Sense and Denotation: Alternative (1)'. *Noûs* **27**: 141–57.

Clark, Michael J., and David Liggins (2012). 'Recent Work on Grounding'. *Analysis* **72**: 812–23.

Colyvan, Mark (2001). *The Indispensability of Mathematics.* Oxford: Oxford University Press.

Congrès International de Philosophie Scientifique, ed. (1936). *Actes du Congrès International de Philosophie Scientifique VII: Logique*. Paris: Hermann.

Cook, Roy (2003). 'Iteration One More Time'. *Notre Dame Journal of Formal Logic* **44**: 63–92.

———————— (2009). 'Hume's Big Brother: Counting Concepts and the Bad Company Objection'. *Synthese* **170**: 349–69.

———————— (2010). 'Conservativeness, Stability, and Abstraction'. *British Journal for the Philosophy of Science* **63**: 673–96.

———————— and Philip A. Ebert (2004). Review of Fine (2002). *British Journal for the Philosophy of Science* **55**: 791–800.

————————, ———————— (2005). 'Abstraction and Identity'. *Dialectica* **59**: 121–39.

Copi, Irving M. (1971). *The Theory of Logical Types.* London: Routledge & Kegan Paul.

Correia, Fabrice (2005). *Existential Dependence and Cognate Notions.* Munich: Philosophia Verlag.

———————— (2008). 'Ontological Dependence'. *Philosophy Compass* **3**: 1013–32.

———————— (2010). 'Grounding and Truth-Functions'. *Logique et analyse* **53/211**: 251–79.

———————— (2014). 'Logical Grounds'. *Review of Symbolic Logic* **7**: 31–59.

———————— and Benjamin Schnieder (2012a). 'Grounding: An Opinionated Introduction'. In Correia and Schnieder (2012b): 1–36.

————————, ————————, eds. (2012b). *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press.

Cotnoir, A.J. (2015). 'Abelian Mereology'. *Logic and Logical Philosophy* **24**: 429–47.

Crossley, J. N., and Michael Dummett, eds. (1965). *Formal Systems and Recursive Functions: Proceedings of the Eighth Logic Colloquium, Oxford.* Amsterdam: North-Holland.

———————— and Lloyd Humberstone (1977). 'The Logic of "Actually"'. *Reports on Mathematical Logic* **8**: 11–29.

Czermak, Johannes, ed. (1993). *Philosophy of Mathematics: Proceedings of the Fifteenth International Wittgenstein Symposium, Part I.* Vienna: Holder–Pichler–Tempsky.

Davis, Robert L. (1953). 'The Number of Structures of Finite Relations'. *Proceedings of the American Mathematical Society* **4**: 486–95.

Diehl, Catharine (ms). 'Fineness of Grain and Grounding: Towards a Lewisian Account'. Unpublished.

Donaldson, Thomas (ms). 'The (Metaphysical) Foundations of Arithmetic?' Forthcoming, *Noûs.*

Dorr, Cian (2016). 'To Be F is to Be G'. *Philosophical Perspectives* **30**: 39–134.

Drake, Frank R. (1974). *Set Theory: An Introduction to Large Cardinals*. Amsterdam: North-Holland.

Dummett, Michael (1963). 'The Philosophical Sentence of Gödel's Theorems'. *Ratio* **5**: 140–55. Rpt. in Dummett (1978): 186–201.*

——————— (1978). *Truth and Other Enigmas.* Cambridge, Mass.: Harvard University Press.

——————— (1981). *Frege's Philosophy of Language.* 2nd edn. Cambridge, Mass.: Harvard University Press.

——————— (1991). *Frege's Philosophy of Mathematics.* Cambridge, Mass.: Harvard University Press.

——————— (1993a). 'Introductory Remarks'. In Czermak (1993): 69–76.

——————— (1993b). *The Seas of Language*. Oxford: Oxford University Press.

——————— (1993c). 'What Is Mathematics About?' In George (1994): 11–26. Rpt. in Dummett (1993a): 429–45.*

——————— (1994). 'Chairman's Address: Basic Law V'. *Proceedings of the Aristotelian Society* **94**: 243–51.

Ebbinghaus, Heinz-Dieter, and Volker Peckhaus (2007). *Ernst Zermelo: An Approach to His Life and Work*. Berlin: Springer.

Ebert, Philip A., and Marcus Rossberg (2016a). 'Introduction to Abstractionism'. In Ebert and Rossberg (2016b): 3–33.

————————, eds. (2016b). *Abstractionism: Essays in the Philosophy of Mathematics.* Oxford: Oxford University Press.

Eklund, Matti (2009). 'Bad Company and Neo-Fregean Philosophy.' *Synthese* **170**: 393–414.

Ellis, Brian (2001). *Scientific Essentialism.* Cambridge: Cambridge University Press.

Enderton, Herbert B. (1977). *Elements of Set Theory.* New York: Academic Press.

Evans, Gareth (1982). *The Varieties of Reference.* Ed. John McDowell. Oxford: Clarendon Press.

Fairchild, Maegan (ms). 'Hybrid Contingentism'. Unpublished.

Feferman, Solomon (1967). 'Set-Theoretical Foundations of Category Theory'. (With an appendix by Georg Kreisel.) In Mac Lane (1969): 201–47.

———————— (1977). 'Categorical Foundations and Foundations of Category Theory'. In Butts and Hintikka (1977): 149–69.

———————— (1984). 'Toward Useful Type-Free Theories I'. *Journal of Symbolic Logic* **49**: 75–111.

———————— (1989). *The Number Systems: Foundations of Algebra and Analysis.* 2nd edn. Providence, R.I.: AMS/Chelsea Publishing.

———————— (2005). 'Predicativity'. In Shapiro (2005): 590–624.

————————, Georg Kreisel, and Steven Orey (1962). '1-Consistency and Faithful Interpretations'. *Archiv für Mathematische Logik und Grundlagenforschung* **5**: 52–63.

Ferreira, Fernando, and Kai F. Wehmeier (2002). 'On the Consistency of the Δ^1_1–CA Fragment of Frege's *Grundgesetze*'. *Journal of Philosophical Logic* **31**: 301–11.

Field, Hartry (1974). 'Quine and the Correspondence Theory'. *Philosophical Review* **83**: 220–228.

———————— (1980). *Science without Numbers: A Defence of Nominaism.* Oxford: Basil Blackwell.

———————— (1984). 'Is Mathematical Knowledge Just Logical Knowledge?' *Philosophical Review* **93**: 509–52. Rpt. with postscript in Field (1989): 79–124.

———————— (1989). *Realism, Mathematics, and Modality.* Oxford: Basil Blackwell.

Fine, Kit (1977). 'Prior on the Construction of Possible Worlds and Instants'. In Prior and Fine (1977): 116–68.

——————— (1981). 'First-Order Modal Theories I: Sets'. *Noûs* **15**: 177–205.

——————— (1994a). 'Essence and Modality'. *Philosophical Perspectives* **8**: 1–16.

——————— (1994b). 'Senses of Essence'. In Sinnott-Armstrong, Raffman, and Asher (1994): 53–73.

——————— (1995a). 'The Logic of Essence'. *Journal of Philosophical Logic* **24**: 241–73.

——————— (1995b). 'Ontological Dependence'. *Proceedings of the Aristotelian Society* **95**: 269–90.

——————— (1998). 'Cantorian Abstraction: A Reconstruction and Defense'. *Journal of Philosophy* **95**: 599–634.

——————— (2002). *The Limits of Abstraction*. Oxford: Clarendon Press.

——————— (2010a). 'Some Puzzles of Ground'. *Notre Dame Journal of Formal Logic* **51**: 97–118.

——————— (2010b). 'Towards a Theory of Part'. *Journal of Philosophy* **107**: 559–89.

——————— (2012a). 'The Pure Logic of Ground'. *Review of Symbolic Logic* **5**: 1–25.

——————— (2012b). 'Guide to Ground'. In Correia and Schnieder (2012b): 37–80.

——————— (2015). 'Unified Foundations for Essence and Ground'. *Journal of the American Philosophical Association* **1**: 296–311.

Fraenkel, Adolf (1919). *Einleitung in die Mengenlehre: Ein gemeinverständliche Einführung in das Reich der unendlichen Größen.* Berlin: Julius Springer.

——————— (1966). *Abstract Set Theory.* 3rd edn. Amsterdam: North-Holland.

Frege, Gottlob (1884). *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung über den Begriff der Zahl.* Breslau: W. Koebner. Rpt. as Frege (1986).* Trans. as Frege (1980). (Cited by section.)

——————— (1891). *Funktion und Begriff*. Jena: Hermann Pohle. Rpt. in Frege (1967): 2–31.* Trans. in Frege (1960): 21–41. (Cited by marginal pagination.)

——————— (1893–1903). *Grundgesetze der Arithmetik*. 2 vols. Jena: Verlag von Hermann Pohl. Rpt. ed. Christian Thiel. Hildesheim: Olms, 2009.*

———————— (1914). 'Logik in der Mathematik'. In Frege (1969): 219–70.* Trans. in Frege (1979): 203–50.

———————— (1960). *Philosophical Writings of Gottlob Frege.* Trans. Peter Geach and Max Black. 2nd edn. Oxford: Blackwell.

———————— (1967). *Kleine Schriften*. Ed. Ignacio Angelelli. Darmstadt: Wissenschaftliche Buchgesellschaft.

———————— (1969). *Nachgelassene Schriften*. Ed. Hans Hermes, Friedrich Kambartel, and Friederich Kaulbach. Hamburg: Meiner.

———————— (1979). *Posthumous Writings.* Trans. Peter Long and Roger White. Oxford: Blackwell.

———————— (1980). *The Foundations of Arithmetic.* Trans. J.L. Austin. 2nd edn. Oxford: Blackwell.

Friedman, Harvey (1971). 'Higher Set Theory and Mathematical Practice'. *Annals of Mathematical Logic* **2**: 325–57.

———————— (2000). 'Normal Mathematics Will Need New Axioms'. *Bulletin of Symbolic Logic* **6**: 401–446.

———————— (2004). 'Note on Faithful Representation in Set Theory without Atoms'. <http://www.cs.nyu.edu/pipermail/fom/2004-January/007845.html>.

Fritz, Peter (ms 1). 'Higher-Order Contingentism, Part 2: Patterns of Indistinguishability'. Forthcoming, *Journal of Philosophical Logic.*

———————— (ms 2). 'Higher-Order Contingentism, Part 3: Expressive Limitations'. Forthcoming, *Journal of Philosophical Logic.*

————————, and Jeremy Goodman (2016). 'Higher-Order Contingentism, Part 1: Closure and Generation'. *Journal of Philosophical Logic* **45**: 645–95.

Gabbay, Dov, and Frank M. Guenther, eds. (1983–87). *Handbook of Philosophical Logic*. Dordrecht: Reidel.

Gallin, Daniel (1975). *Intensional and Higher-Order Modal Logic*. Amsterdam: North-Holland.

García-Carpintero, Manuel, and Josep Macià, eds. (2006). *Two-Dimensional Semantics.* Oxford: Oxford University Press.

George, Alexander, ed. (1994). *Mathematics and Mind.* Oxford: Oxford University Press.

Gödel, Kurt (1964). 'What is Cantor's Continuum Problem?' In Benacerraf and Putnam (1964): 258–73. Rpt. in Gödel (1986–2003): **2**.254–70.*

———————— (1986–2003). *Collected Works.* Ed. Solomon Feferman et al. 5 vols. Oxford: Oxford University Press.

Goodman, Jeremy (2016). 'Williamson on Necessitism'. *Canadian Journal of Philosophy* **46.4–5**: 613–39.

———————— (2017). 'Reality is Not Structured'. *Analysis* **77**: 43–53.

———————— (ms). 'Theories of Aboutness'. Unpublished.

Goodman, Nelson, and Willard Van Orman Quine (1947). 'Steps toward a Constructive Nominalism'. *Journal of Symbolic Logic* **12**: 105–22.

Grossman, Reinhardt (1983). *The Categorical Structure of the World.* Bloomington, Ind.: Indiana University Press.

Halbach, Volker (2014). *Axiomatic Theories of Truth.* Rev. edn. Cambridge: Cambridge University Press.

————————, and Horsten, Leon [2005]. 'Computational Structuralism'. *Philosophia Mathematica* (3rd ser.) **13**: 174–86.

Hale, Bob (1987). *Abstract Objects.* Oxford: Blackwell.

———————— (1997). *'Grundlagen* §64'. *Proceedings of the Aristotelian Society* **97**: 243–61. Rpt. in Hale and Wright (2001b): 91–116.*

———————— (2001). 'A Response to Potter and Smiley: Abstraction by Recarving'. *Proceedings of the Aristotelian Society* **101**: 339–58.

————————, and Aviv Hoffman, eds. (2010). *Modality: Metaphysics, Logic, and Epistemology.* Oxford: Oxford University Press.

————————, and Crispin Wright (2000). 'Implicit Definition and the A Priori'. In Boghossian and Peacocke (2000): 286–319. Rpt. in Hale and Wright (2001b): 117–

50.*

———————, ——————— (2001a). 'To Bury Caesar . . . '. In Hale and Wright (2001b): 335–96.

———————, ——————— (2001b). *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics.* Oxford: Oxford University Press.

Hallett, Michael (1984). *Cantorian Set Theory and Limitation of Size.* Oxford: Clarendon Press.

Hamkins, Joel David (2012). 'The Set-Theoretic Multiverse'. *Review of Symbolic Logic* 5: 416–99.

——————— (2015). 'Is the Dream Solution to the Continuum Hypothesis Attainable?' *Notre Dame Journal of Formal Logic* **56**: 135–45.

———————, and Øystein Linnebo (ms). 'The Modal Logic of Set-Theoretic Potentialism and the Potentialist Maximality Principles'. Unpublished; arXiv:1708.01644.

Hardy, G.H., and E.M. Wright (1960). *An Introduction to the Theory of Numbers.* 4th edn. Oxford: Clarendon Press.

Hazen, A.P. (1985). Review of Wright (1983). *Australasian Journal of Philosophy* **63**: 251–54.

——————— (1986). 'Logical Objects and the Paradox of Burali-Forti'. *Erkenntnis* **24**: 283–91.

——————— (1993). 'Against Pluralism'. *Australasian Journal of Philosophy* **71**: 132–45.

——————— (1996). 'Worlds as Complete Novels'. *Analysis* **56**: 33–38.

——————— (1997a). 'Relations in Lewis's Framework without Atoms'. *Analysis* **57**: 243–48.

——————— (1997b). 'Relations in Monadic Third-Order Logic'. *Journal of Philosophical Logic* **26**: 619–628.

——————— (2004). 'Hypergunk'. *Monist* **87**: 322–38.

Heck, Richard (2011). *Frege's Theorem.* Oxford: Oxford University Press.

——————— (2012). *Reading Frege's 'Grundgesetze'.* Oxford: Oxford University Press.

————————, ed. (1997). *Logic, Language, and Thought: Essays in Honour of Michael Dummett.* Oxford: Clarendon Press.

Hellman, Geoffrey (1989). *Mathematics without Numbers.* Oxford: Clarendon Press.

———————— (2001). 'Three Varieties of Mathematical Structuralism'. *Philosophia Mathematica* (3rd ser.) **9**: 184–211.

———————— (2003). 'Does Category Theory Provide a Framework for Mathematical Structuralism?' *Philosophia Mathematica* (3rd ser.) **11**: 129–57.

Henle, Paul, Horace M. Kallen, and Suzanne K. Langer, eds. (1951). *Structure, Method, and Meaning: Essays in Honor of Henry M. Scheffer*. New York: Liberal Arts Press.

Herstein, I.M. (1975). *Topics in Algebra*. 2nd edn. New York: Wiley.

Hilbert, David (1926). 'Über das Unendliche'. *Mathematische Annalen* **95**: 161–90. Trans. as Hilbert (1967).

———————— (1967). 'On the Infinite'. In Van Heijenoort (1967): 367–92.

Hindley, J. Roger, and Jonathan P. Seldin (2008). *Lambda Calculus and Combinators.* 2nd edn. Cambridge: Cambridge University Press.

Hodes, Harold (1984). 'Logicism and the Ontological Commitments of Arithmetic'. *Journal of Philosophy* **81**: 123–49.

———————— (2015). 'Why Ramify?' *Notre Dame Journal of Formal Logic* **56**: 379–415.

Hoeltje, Miguel, Benjamin Schnieder, and Alex Steinberg, eds. (2013). *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*. Munich: Philosophia.

Hoffman, Joshua, and Gary S. Rosenkrantz (1994). *Substance among Other Categories.* Cambridge: Cambridge University Press.

Horsten, Leon (2005). 'Canonical Naming Systems'. *Minds and Machines* **15**: 229–57.

Hovda, Paul (2009). 'What Is Classical Mereology?' *Journal of Philosophical Logic* **38**: 55–82.

Husserl, Edmund (1891). *Die Philosophie der Arithmetik: Psychologische und logische Untersuchungen.* Halle: Pfeffer. Rpt. in Husserl (1950–) **12**. Trans. Dallas Willard

as Husserl (1980–) 10.

———— (1950–). *Husserliana: Gesammelte Werke*. Various editors. The Hague: Nijhoff.

———— (1980–). *Edmund Husserl: Collected Works*. Various editors. Dordrecht: Kluwer.

Incurvati, Luca (2012). 'How to Be a Minimalist about Sets'. *Philosophical Studies* 159: 69–87.

Isaacson, Daniel (2011). 'The Reality of Mathematics and the Case of Set Theory'. In Novák and Simonyi (2011): 1–76.

Jané, Ignacio, and Gabriel Uzquiano (2004). 'Well- and Non-Well-Founded Fregean Extensions'. *Journal of Philosophical Logic* 33: 437–65.

Jech, Thomas (2002). *Set Theory*. Rev. (3rd Millennium) edn. Berlin: Springer.

Kalderon, Mark Eli, ed. (2002). *Fictionalism in Metaphysics*. Oxford: Oxford University Press.

Kamareddine, Fairouz, Twan Laan, and Rob Nederpelt (2004). *A Modern Perspective on Type Theory: From Its Origins until Today*. Dordrecht: Kluwer.

Kanamori, Akihiro (2003). *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings*. 2nd edn. Berlin: Springer.

———— (2012). 'In Praise of Replacement'. *Bulletin of Symbolic Logic* 18: 46–90.

Kaplan, David (1964). *Foundations of Intensional Logic*. Ph.D. thesis, University of California, Los Angeles.

———— (1968). 'Quantifying In'. *Synthese* 19: 178–214.

———— (1975). 'How to Russell a Frege-Church'. *Journal of Philosophy* 72: 716–29.

———— (1978). 'Dthat'. *Syntax and Semantics* 9: 221–43.

———— (1989). 'Demonstratives'. In Almog, Perry, and Wettstein (1989): 481–563.

Kennedy, Juliette, and Roman Kossak, eds. (2011). *Set Theory, Arithmetic, and Foundations of Mathematics: Theorems, Philosophies*. Cambridge: Cambridge University

Press.

Keränen, Jukka (2001). 'The Identity Problem for Realist Structuralism'. *Philosophia Mathematica* (3rd ser.) **9**: 308–30.

———————— (2006). 'The Identity Problem for Realist Structuralism II: A Reply to Shapiro'. In MacBride (2006): 146–63.

Ketland, Jeffrey (2002). 'Hume = Small Hume'. *Analysis* **62**: 92–93.

———————— (ms). 'Abstract Structure'. Typescript, January 2015.

Kitcher, Philip (1983). *The Nature of Mathematical Knowledge.* Oxford: Oxford University Press.

Klement, Kevin (2002). *Frege and the Logic of Sense and Reference.* New York: Routledge.

———————— (2003). 'The Number of Senses'. *Erkenntnis* **58**: 303–23.

———————— (2010). 'The Senses of Functions in the Logic of Sense and Denotation'. *Bulletin of Symbolic Logic* **16**: 153–88.

Koellner, Peter (2003). *The Search for New Axioms*. Ph.D. thesis, Massachusetts Institute of Technology.

———————— (2009). 'On Reflection Principles'. *Annals of Pure and Applied Logic* **157**: 206–19.

Koslicki, Kathrin (2008). *The Structure of Objects.* Oxford: Oxford University Press.

———————— (2012). 'Varieties of Ontological Dependence'. In Correia and Schneider (2012b): 186–213.

Kouri, Teresa (2016). *'Ante Rem* Structuralism and the No-Naming Constraint'. *Philosophia Mathematica* (3rd ser.) **24**: 117–28.

Krämer, Stephan (2014). *On What There Is for Things to Be: Ontological Commitment and Second-Order Quantification*. Frankfurt: Klostermann.

Kreisel, Georg (1965). 'Mathematical Logic'. In Saaty (1964–65): 3.95–195.

———————— (1967). 'Informal Rigour and Consistency Proofs'. In Lakatos (1967): 138– 16.

———————— and Jean-Louis Krivine (1966). *Élements de logique mathématique: théorie des modèles*. Paris: Dunod.

————————, ———————— (1967). *Elements of Mathematical Logic: Model Theory.* Amsterdam: North-Holland. Eng. trans. of Kreisel and Krivine (1966).

Kripke, Saul (1976). 'Is There a Problem about Substitutional Quantification?' In Evans and McDowell (1976): 325–419.

———————— (1979). 'A Puzzle about Belief'. In Margalit (1979): 239–83.

———————— (1992). 'Logicism, Wittgenstein, and De Re Beliefs About Numbers'. Whitehead Lectures, Harvard University. Unpublished.

Kunen, Kenneth (1980). *Set Theory: An Introduction to Independence Proofs.* Amsterdam: North-Holland.

———————— (2013). *Set Theory.* Rev. edn. London: College Publications.

Ladyman, James (1998). 'What is Structural Realism?' *Studies in History and Philosophy of Science* **29**: 409–24.

———————— (2007). 'On the Identity and Diversity of Individuals'. *Proceedings of the Aristotelian Society Supplement* **81**: 23–43.

Lakatos, Irme, ed. (1967). *Problems in the Philosophy of Mathematics.* Amsterdam: North-Holland.

Lawvere, F. William (1964). 'An Elementary Theory of the Category of Sets'. *Proceedings of the National Academy of Sciences* **52**: 1506–11.

Lear, Jonathan (1977). 'Sets and Semantics'. *Journal of Philosophy* **74**: 86–102.

Leitgeb, Hannes, and James Ladyman (2008). 'Criteria of Identity and Structuralist Ontology'. *Philosophia Mathematica* (3rd ser.) **16**: 388–96.

Lévy, Azriel (1960a). 'Axiom Schemata of Strong Infinity in Set Theory'. *Pacific Journal of Mathematics* **10**: 223–28.

———————— (1960b). 'Principles of Reflection in Axiomatic Set Theory'. *Fundamenta Mathematicae* **49**: 1–10.

Lewis, David (1983). 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy* **61**: 343–77. Rpt. in Lewis (1999): 8–55.*

———————— (1983–86). *Philosophical Papers.* 2 vols. Oxford: Oxford University Press.

———————— (1984). 'Putnam's Paradox'. *Australasian Journal of Philosophy* **62**: 221–36. Rpt. in Lewis (1999): 56–77.*

———————— (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

———————— (1990). 'Noneism or Allism?' *Mind* **99**: 24–31. Rpt. in Lewis (1999): 152–63.*

———————— (1991). *Parts of Classes.* Oxford: Blackwell.

———————— (1999). *Papers in Metaphysics and Epistemology.* Cambridge: Cambridge University Press.

Linnebo, Øystein (2008). 'Structuralism and the Notion of Dependence'. *Philosophical Quarterly* **58**: 59–79.

———————— (2009). 'The Individuation of the Natural Numbers'. In Bueno and Linnebo (2009): 220–38.

———————— (2013). 'The Potential Hierarchy of Sets'. *Review of Symbolic Logic* **6**: 205–228.

———————— (2016). 'Impredicativity in the Neo-Fregean Program'. In Ebert and Rossberg (2016b): 247–68.

————————, and Richard Pettigrew (2011). 'Category Theory as an Autonomous Foundation'. *Philosophia Mathematica* (3rd ser.) **19**: 227–54.

————————, ———————— (2014). 'Two Types of Abstraction for Structuralism'. *Philosophical Quarterly* **64**: 267–83.

———————— and Agustín Rayo (2012). 'Hierarchies Ontological and Ideological'. *Mind* **121**: 269–308.

Linsky, Leonard, ed. (1952). *Semantics and the Philosophy of Language: A Collection of Readings.* Urbana, Ill.: University of Illinois Press.

Lowe, E.J. (1989). *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms.* Oxford: Blackwell.

———————— (2006). *The Four-Category Ontology: A Metaphysical Foundation for Natural Science.* Oxford: Clarendon Press.

Mac Lane, Saunders (1986). *Mathematics: Form and Function.* Berlin: Springer.

————————, ed. (1969). *Reports of the Midwest Category Seminar III.* Berlin: Springer.

————————, and Garrett Birkhoff (1988). *Algebra.* 3rd edn. Providence, R.I.: AMS/Chelsea Publishing.

MacBride, Fraser (2003). 'Speaking with Shadows: A Study of Neo-Logicism'. *British Journal for the Philosophy of Science* **54**:103–63.

————————, ed. (2006). *Identity and Modality*. Oxford: Oxford University Press.

McGee, Vann (1996). 'Logical Operations'. *Journal of Philosophical Logic* **25**: 567–80.

———————— (1997). 'How We Learn Mathematical Language'. *Philosophical Review* **106**: 35–68.

———————— (2000). 'Everything'. In Sher and Tieszen (2000): 54–78.

———————— (2006). 'There Are Many Things'. In Thomson and Byrne (2006): 93–122.

McKenzie, Ralph, and Donald Monk (1982). 'Chains in Boolean Algebras'. *Annals of Mathematical Logic* **22**: 137–75.

Maddy, Penelope (1988a). 'Believing the Axioms I'. *Journal of Symbolic Logic* **53**: 481–511.

———————— (1988b). 'Believing the Axioms II'. *Journal of Symbolic Logic* **53**: 736–64.

————————(1990). *Realism in Mathematics.* Oxford: Clarendon Press.

———————— (1997). *Naturalism in Mathematics.* Oxford: Clarendon Press.

Mancosu, Paolo (2017). *Abstraction and Infinity.* Oxford: Oxford University Press.

Margalit, Avishai, ed. (1979). *Meaning and Use: Papers Presented at the Second Jerusalem Philosophical Encounter, April 1976.* Dordrecht: Reidel.

Mates, Benson (1952). 'Synonymity'. In Linsky (1952): 111–36.

Menzel, Christopher (1986). 'On the Iterative Explanation of the Paradoxes'. *Philosophical Studies* **49**: 37–61.

———————— (2014). 'Wide Sets, ZFCU, and the Iterative Conception'. *Journal of Philosophy* **111**: 57–83.

Mirimanoff, Dimitry (1917). 'Les antinomies de Russell et Burali-Forti et le problème fondamental de la théorie des ensembles'. *L'enseignement mathématique* **19**: 37–52.

Montague, Richard (1965). 'Set Theory and Higher-Order Logic'. In Crossley and Dummett (1965): 131–48.

Muskens, Reinhard (1989). 'A Relational Formulation of the Theory of Types'. *Linguistics and Philosophy* **12**: 325–46.

————————(2007). 'Intensional Models for the Theory of Types'. *Journal of Symbolic Logic* **72**: 98–118.

Myhill, John (1974). 'The Undefinability of the Set of Natural Numbers in the Ramified *Principia*'. In Nakhnikian (1974): 19–27.

———————— (1977). 'A Refutation of an Unjustified Attack on the Axiom of Reducibility'. In Roberts (1979): 81–90.

Nakhnikian, George, ed. (1974). *Bertrand Russell's Philosophy*. London: Duckworth.

Nelson, Edward (1986). *Predicative Arithmetic*. Princeton: Princeton University Press.

Novák, Zsolt, and András Simonyi, eds. (2011). *Truth, Reference, and Realism*. Budapest: Central European University Press.

Olson, Eric T. (2007). *The Human Animal: Personal Identity without Psychology*. Oxford: Oxford University Press.

Oppenheimer, Paul E., and Edward Zalta (2011). 'Relations versus Functions at the Foundations of Logic: Type-Theoretic Considerations'. *Journal of Logic and Computation* **21**: 351–74.

Orey, Steven (1959). 'Model Theory for the Higher Order Predicate Calculus'. *Transactions of the American Mathematical Society* **92**: 72–84.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press.

———————— (2012). 'We Are Not Human Beings'. *Philosophy* **87**: 5–28.

Parsons, Charles (1977). 'What is the Iterative Conception of Set?' In Butts and Hintikka (1977): 335–67. Rpt. in Parsons (1983a): 268–97.*

———————— (1983a). *Mathematics in Philosophy: Selected Essays*. Ithaca: Cornell University Press.

———————— (1983b). 'Sets and Modality'. In Parsons (1983a): 298–341.

———————— (1990). 'The Structuralist View of Mathematical Objects'. *Synthese* **84**: 303–46.

———————— (2004). 'Structuralism and Metaphysics'. *Philosophical Quarterly* **54**: 56–77.

———————— (2008). *Mathematical Thought and Its Objects.* Cambridge: Cambridge University Press.

Paseau, Alexander (2001). 'Should the Logic of Set Theory Be Intuitionistic?' *Proceedings of the Aristotelian Society* **101**: 369–78.

———————— (2003). 'The Open-Endedness of the Set Concept and the Semantics of Set Theory'. *Synthese* **135**: 379–99.

———————— (2007). 'Boolos on the Justification of Set Theory'. *Philosophia Mathematica* (3rd ser.) **15**: 30–53.

———————— (2009). 'Reducing Arithmetic to Set Theory'. In Bueno and Linnebo (2009): 35–55.

Pierce, Benjamin C. (2002). *Types and Programming Languages.* Cambridge, Mass.: MIT Press.

Pollard, Stephen (1988). 'Plural Quantification and the Axiom of Choice'. *Philosophical Studies* **54**: 393–97.

———————— (1989). 'Choice Again'. *Philosophical Studies* **66**: 285–96.

Potter, Michael (2004). *Set Theory and Its Philosophy: A Critical Introduction.* Oxford: Oxford University Press.

———————— and Timothy Smiley (2001). 'Abstraction by Recarving'. *Proceedings of the Aristotelian Society* **101**: 327–38.

————————, ———————— (2002). 'Recarving Content: Hale's Final Proposal'. *Proceedings of the Aristotelian Society* **102**: 351–54.

Prior, A.N. (1971). *Objects of Thought.* Ed. P.T. Geach and A.J.P. Kenny. Oxford: Clarendon Press.

————————, and Kit Fine (1977). *Worlds, Times, and Selves.* London: Duckworth.

Putnam, Hilary (1971). *Philosophy of Logic.* London: Allen & Unwin. Rpt. in Putnam (1979): 323–58.

———— (1979). *Mathematics, Matter, and Method: Philosophical Papers, Volume I.* 2nd edn. Cambridge: Cambridge University Press.

Quine, Willard Van Orman (1948). 'On What There Is'. *Review of Metaphysics* **2**: 21–38. Rev. rpt. in Quine (1980): 1–19.*

———— (1960a). 'Carnap and Logical Truth'. *Synthese* **12**: 350–74. Rev. rpt. in Quine (1976): 107–32.*

———— (1960b). *Word and Object.* Cambridge, Mass.: MIT Press.

———— (1972). *The Ways of Paradox and Other Essays.* Rev. edn. Cambridge, Mass.: Harvard University Press.

———— (1980). *From a Logical Point of View: Nine Logico-Philosophical Essays.* 2nd edn. Cambridge, Mass.: Harvard University Press.

———— (1981a). *Theories and Things.* Cambridge, Mass.: Harvard University Press.

———— (1981b). 'Things and their Place in Theories'. In Quine (1981a): 1–23.

———— (1986). *Philosophy of Logic.* 2nd edn. Cambridge, Mass.: Harvard University Press.

Raven, Michael J. (2015). 'Ground'. *Philosophy Compass* **10**: 321–33.

Rayo, Agustín (2002). 'Word and Objects'. *Noûs* **36**: 436–64.

———— (2006). 'Beyond Plurals'. In Rayo and Uzquiano (2006): 220–54.

———— (2007). 'Ontological Commtiment'. *Philosophy Compass* **2**: 428–44.

———— (2013). *The Construction of Logical Space.* Oxford: Oxford University Press.

———— (2016). 'Neo-Fregeanism Reconsidered'. In Ebert and Rossberg (2016b): 203–220.

———— and Gabriel Uzquiano (1999). 'Toward a Theory of Second-Order Consequence'. *Notre Dame Journal of Formal Logic* **40**: 315–25.

————————, ————————, eds. (2006). *Absolute Generality.* Oxford: Oxford University Press.

————————, and Timothy Williamson (2003). 'A Completeness Theorem for Unrestricted First-Order Languages'. In Beall (2003): 331–55.

————————, and Stephen Yablo (2001). 'Nominalism through De-Nominalization'. *Noûs* 35: 74–92.

Räz, Tim (2015). 'Say My Name: An Objection to Ante Rem Structuralism'. *Philosophia Mathematica* (3rd ser.) 23: 116–25.

Reinhardt, W.N. (1974). 'Remarks on Reflection Principles, Large Cardinals, and Elementary Embeddings'. In Scott and Jech (1971–74): 2.189–205.

Roberts, George W., ed. (1979). *Bertrand Russell Memorial Volume*. London: Allen & Unwin.

Rosen, Gideon (1992). *Remarks on Modern Nominalism.* Ph.D. thesis, Princeton University.

———————— (2010). 'Metaphysical Dependence: Grounding and Reduction'. In Hale and Hoffman (2010): 109–36.

———————— and Yablo, Stephen (ms). 'Solving the Caesar Problem—with Metaphysics'. Dated 2006; forthcoming in a *Festschrift* for Crispin Wright.

Rudin, Walter (1976). *Principles of Mathematical Analysis.* 3rd edn. New York: McGraw-Hill.

Rumfitt, Ian (2001). 'Hume's Principle and the Number of All Objects'. *Noûs* 35: 515–41.

———————— (2015). *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford: Oxford University Press, 2015.

Russell, Bertrand (1903). *The Principles of Mathematics.* Cambridge: Cambridge University Press.

———————— (1936). 'The Limits of Empiricism'. *Proceedings of the Aristotelian Society* 36: 131–50.

Saaty, Thomas L., ed. (1964–65). *Lectures on Modern Mathematics.* 3 vols. New York: Wiley.

Salmon, Nathan (2010). 'Lambda in Sentences with Designators: An Ode to Complex Predication'. *Journal of Philosophy* **107**: 445–68.

Schaffer, Jonathan (2010). 'Monism: The Priority of the Whole'. *Philosophical Review* **119**: 31–76.

Schirn, Matthias, ed. (1998). *The Philosophy of Mathematics Today.* Oxford: Clarendon Press.

Schnieder, Benjamin (2011). 'A Logic for "Because"'. *Review of Symbolic Logic* **4**: 445–65.

Schwartzkopff, Robert (2011). 'Numbers as Ontologically Dependent Objects: Hume's Principle Revisited'. *Grazer Philosophische Studien* **81**: 353–74.

Scott, Dana S. (1955). 'Definitions by Abstraction in Axiomatic Set Theory'. *Bulletin of the American Mathematical Society* **61**: 442.

——————— (1971). 'On Engendering an Illusion of Understanding'. *Journal of Philosophy* **68**: 787–807.

——————— (1974). 'Axiomatizing Set Theory'. In Scott and Jech (1971–74): **2**.207–14.

———————, and Jech, Thomas J., ed. (1971–74). *Axiomatic Set Theory.* 2 vols. Providence, R.I.: American Mathematical Society.

Shapiro, Stewart (1982). 'Acceptable Notation'. *Notre Dame Journal of Formal Logic* **23**: 14–20.

——————— (1987). 'Principles of Reflection and Second-Order Logic'. *Journal of Philosophical Logic* **16**: 309–33.

——————— (1991). *Foundations without Foundationalism: A Case for Second-Order Logic.* Oxford: Oxford University Press.

——————— (1997). *Philosophy of Mathematics: Structure and Ontology.* Oxford: Oxford University Press.

——————— (2006). 'Structure and Identity'. In MacBride (2006): 109–45.

————————, ed. (2005). *The Oxford Handbook of Philosophy of Mathematics and Logic.* Oxford: Oxford University Press.

———————— and Gabriel Uzquiano (2008). 'Frege Meets Zermelo: A Perspective on Ineffability and Reflection'. *Review of Symbolic Logic* 1: 241–266.

———————— and Alan Weir (1999). 'New V, ZF, and Abstraction'. *Philosophia Mathematica* (3rd ser.) 7: 293–321.

Sher, Gila (1991). *The Bounds of Logic: a Generalized Viewpoint*. Cambridge, Mass.: MIT Press.

————————, and Richard Tieszen, eds. (2000). *Between Logic and Intuition: Essays in Honor of Charles Parsons.* Cambridge: Cambridge University Press.

Shoenfield, Joseph R. (1967). *Mathematical Logic*. Reading, Mass.: Addison-Wesley.

———————— (1977). 'Axioms of Set Theory'. In Barwise (1977): 321–44.

Simons, Peter (1987). *Parts: A Study in Ontology.* Oxford: Clarendon Press.

Sinnott-Armstrong, Walter, Diana Raffman, and Nicholas Asher, eds. (1994). *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press.

Smith, Peter (2013). *An Introduction to Gödel's Theorems.* 2nd edn. Cambridge: Cambridge University Press.

Solomon, Graham (1989). 'What Became of Russell's "Relation Arithmetic"?' *Russell* 9: 168–73.

Sorensen, Roy A. (1988). *Blindspots.* Oxford: Clarendon Press.

Steiner, Mark (1975). *Mathematical Knowledge.* Ithaca: Cornell University Press.

———————— (2011). 'Kripke on Logicism, Wittgenstein, and *De Re* Beliefs about Numbers'. In Berger (2011): 160–76.

Studd, J.P. (2013). 'The Iterative Conception of Set: A (Bi-)Modal Axiomatisation'. *Journal of Philosophical Logic* 42: 697–725.

Sullivan, Peter, and Michael Potter (1997). 'Hale on Caesar'. *Philosophia Mathematica* (3rd ser.) 5: 135–52.

Tait, William (1998). 'Zermelo's Conception of Set Theory and Reflection Principles'. In Schirn (1998): 469–83.

Takeuti, Gaesi (1954). 'Construction of the Set Theory from the Theory of Ordinal Numbers'. *Journal of the Mathematical Society of Japan* 6: 196–220.

——————— (1965). 'A Formalization of the Theory of Ordinal Numbers'. *Journal of Symbolic Logic* 30: 295–317.

Tarski, Alfred (1935). 'Der Wahrheitsbegriff in den formalisierten Sprachen'. *Studia Philosophica* 1: 261–405. Trans. as Tarski (1983a).

——————— (1936a). 'Über den Begriff der logischen Folgerung'. In Congrès International de Philosophie Scientifique (1936): 1–11. Trans. as Tarski (1983c).

——————— (1956). *Ordinal Algebras.* With appendices by Chen-Chung Chang and Bjarni Jónsson. Amsterdam: North-Holland.

——————— (1983a). 'The Concept of Truth in Formalized Languages'. In Tarski (1983b): 152–278.

——————— (1983b). *Logic, Semantics, Metamathematics.* Trans. J.H. Woodger; ed. John Corcoran. 2nd edn. Indianapolis: Hackett.

——————— (1983c). 'On the Concept of Logical Consequence'. In Tarski (1983b): 409–20.

——————— (1986). 'What are Logical Notions?' *History and Philosophy of Logic* 7: 143–154.

———————, Andrzej Mostowski, and Raphael M. Robinson (1953). *Undecidable Theories.* Amsterdam: North-Holland.

Thomson, Judith Jarvis, and Alex Byrne, eds. (2006). *Content and Modality: Themes from the Philosophy of Robert Stalnaker.* Oxford: Oxford University Press.

Trogdon, Kelly (2013). 'An Introduction to Grounding'. In Hoeltje, Schnieder, and Steinberg (2013): 97–122.

Uzquiano, Gabriel (2003). 'Plural Quantification and Classes'. *Philosophia Mathematica* (3rd ser.) 11: 67–81.

———————— (2015). 'Varieties of Indefinite Extensibility'. *Notre Dame Journal of Formal Logic* **56**: 147–66.

Van Aken, James (1986). 'Axioms for the Set-Theoretic Hierarchy'. *Journal of Symbolic Logic* **51**: 992–1004.

Van Heijenoort, Jean, ed. (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931.* Cambridge, Mass.: Harvard University Press.

Varzi, Achille (2007). 'Supervaluationism and Its Logics'. *Mind* **116**: 663–76.

———————— (2016). 'Mereology'. Stanford Encyclopedia of Philosophy, Winter 2016 edn. <https://plato.stanford.edu/archives/win2016/entries/mereology/>.

Von Neumann, John (1923). 'Zur Einführung der transfiniten Zahlen'. *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae: Sectio Scientiarum Mathematicarum* **1**: 199–208.

———————— (1925). 'Eine Axiomatisierung der Mengenlehre'. *Journal für die reine und angewandte Mathematik* **154**: 219–40. Trans. in Van Heijenoort [1967]: 393–413.

Walsh, Sean (2016). 'Predicativity, the Russell-Myhill Paradox, and Church's Intensional Logic'. *Journal of Philosophical Logic* **45**: 277–326.

Walton, Kendall (1990). *Mimesis as Make Believe: On the Foundations of the Representational Arts.* Cambridge, Mass.: Harvard University Press.

Wang, Hao (1977). 'Large Sets'. In Butts and Hintikka (1977): 309–33.

Welch, Philip (2012). 'Global Reflection Principles'. Isaac Newton Institute Preprint Series. NI12051-SAS.

————————, and Leon Horsten (2016). 'Reflecting on Absolute Infinity'. *Journal of Philosophy* **113**: 89–111.

Westerhoff, Jan (2005). *Ontological Categories: Their Nature and Significance*. Oxford: Clarendon Press.

Wetzel, Linda (1989). 'That Numbers Could Be Objects'. *Philosophical Studies* **56**: 273–92.

Whitehead, A.N., and Bertrand Russell (1927). *Principia Mathematica.* 2nd edn. Cambridge: Cambridge University Press.

Williamson, Timothy (1994). *Vagueness.* London: Routledge.

———— (1998). 'Indefinite Extensibility'. *Grazer philosophische Studien* **55**: 1–24.

———— (1999). 'Existence and Contingency'. *Proceedings of the Aristotelian Society (Supplement)* **73**: 181–203; corrected rpt. as Williamson (2000a).

———— (2000a). 'Existence and Contingency'. *Proceedings of the Aristotelian Society* **100**: 321–43.

———— (2000b). *Knowledge and Its Limits.* Oxford: Oxford University Press.

———— (2003). 'Everything'. *Philosophical Perspectives* **17**: 415–65.

———— (2007). *The Philosophy of Philosophy.* Oxford: Blackwell.

———— (2013). *Modal Logic as Metaphysics.* Oxford: Oxford University Press.

———— (2016). 'Abductive Philosophy'. *Philosophical Forum* **47**: 263–80.

———— (2017). 'Semantic Paradoxes and Abductive Methodology'. In Armour-Garb (2017): 325–46.

Woodin, W. Hugh (2001a). 'The Continuum Hypothesis: Part I.' *Notices of the American Mathematical Society* **48**: 567–76.

———— (2001b). 'The Continuum Hypothesis: Part II.' *Notices of the American Mathematical Society* **48**: 682–90.

———— (2011). 'The Continuum Hypothesis, the Generic Multiverse of Sets, and the Ω Conjecture'. In Kennedy and Kossak (2011): 13–42.

Wright, Crispin (1983). *Frege's Conception of Numbers as Objects.* Aberdeen: Aberdeen University Press.

———— (1997). 'On the Philosophical Significance of Frege's Theorem'. In Heck (1997): 201–44. Rpt. in Hale and Wright (2001b): 272–306.*

———— (1999). 'Is Hume's Principle Analytic?' *Notre Dame Journal of Formal Logic* **40**: 6–30. Rpt. in Hale and Wright (2001b): 307–332.*

Yablo, Stephen (2005). 'The Myth of the Seven'. In Kalderon (2005): 88–117.

———— (2008). 'Carving Content at the Joints'. *Canadian Journal of Philosophy Supplement* **38**: 145–77.

Zalta, Edward (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics.* Dordrecht: Reidel.

——————— (1988). *Intensional Logic and the Metaphysics of Intentionality*. Cambridge, Mass.: MIT Press.

Zermelo, Ernst (1930). 'Über Grenzzahlen und Mengenbereiche'. *Fundamenta Mathematicae* 16: 29–47. Rpt. and trans. in Zermelo (2010–13): 1.400–31.*

——————— (2010–13). *Collected Works/Gesammelte Werke.* Ed. Heinz-Dieter Ebbinghaus, Craig G. Fraser, and Akihiro Kanamori. 2 vols. Berlin: Springer.