

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Philosophy

2016

The Metaphysical Basis of Logic

McSweeney, Michaela McSweeney

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01tb09j8145>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

THE METAPHYSICAL BASIS OF LOGIC

Michaela Markham McSweeney

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

PHILOSOPHY

Adviser: Shamik Dasgupta

September 2016

© Copyright by Michaela Markham McSweeney, 2016. All rights reserved.

Dedication and Acknowledgements

This dissertation is dedicated to Nelson P. Lande, who allowed me to ask way too many philosophical questions in his introductory logic course (many of which I am still grappling with today), who taught me to love philosophy, and logic, on my own terms, and who is the greatest teacher I have ever had.

I'm incredibly grateful to my adviser, Shamik Dasgupta, so much so that the gratefulness is probably ineffable, but I'll try to say something anyway because that seems to be what I do. So thank you, Shamik, for finding the perfect balance between giving me the autonomy to figure out what I wanted to do and say on my own, and always pushing me to make it better. And for having the philosophical insight to help me actually succeed in making it better. For advising so well, both personally and academically, someone who is likely very nearly impossible to advise, who is stubborn and terrible at dealing with authority. Somehow we both survived the process and (I hope) don't even hate each other, and that alone deserves some sort of medal. For being such a tremendous philosophical role model. And most importantly, for believing in my project and the way I want to do philosophy, which is big and weird (I'd like to think in a good way) and hopefully deep but definitely flawed (though I'd like to think not deeply flawed) and not contained in any way to this dissertation.

I also especially want to thank Hans Halvorson, Gideon Rosen, Jonathan Schaffer, Theodore Sider, and Jason Turner. Hans for being one of the best teachers I've ever had, and for obviously being a huge influence, and help, philosophically. But above all for his endless support, empathy, thoughtfulness, loyalty, and friendship, that I couldn't have survived graduate school without. Gideon for his countless sets of comments, so often delivered so quickly after I sent him a draft that I have a feeling he might be superhuman, for setting my head straight when I needed it reset, for tearing my arguments down and helping me build them back up. Jonathan for his unwavering support, for his deep and careful thoughts about my work, and for being so unrepentantly kind. Ted for believing in me from the beginning, for pushing for me, for making it possible for me to get where I am, and for taking me totally seriously, for always talking to me as an equal, and for looking out for me. And, of course, for his work, which has been such a huge influence on me. Jason, for always being willing to make time for me, for being incredibly patient and never patronizing, for being willing to talk technical minutia for far longer than anyone ought to be willing, and for helping me think through so much. And, sort of, for the difficult-to-deal-with objections.

Special thanks to Kathrin Koslicki, for being my hero and role model in pretty much every way. Elizabeth Barnes, for, among many things, all the behind-the-scenes support. Dan Korman, for the countless times you've given me advice and help and friendship. Raul Saucedo and June Gruber, for being there for me at a really hard time, and for being so beautiful. Rebecca Chan, for being my oldest and best work buddy, and an amazing friend. Dan Kaufman, for loyalty, support, understanding, and weirdness. Sukaina Hirji and Daniel Wodak, for being my Princeton family. To Katie Elliott, Elizabeth Miller, Erica Shumener, Elanor Taylor. There are really no words I could say about these four women that would adequately convey how much they've done for me and how much they mean to me. Ole Thomasson Hjortland, for being a dreamer. And to Mark Balaguer, for consistently reminding me what is really important.

I have many other people to thank for help, directly and indirectly, with my dissertation: Boris Alexeev, Yuval Avnir, Andrew Bailey, David Baker, Sophie Ban, David Barnett, Thomas

William Barrett, Nancy Bauer, Megan Bender, Sara Bernstein, Harjit Bhogal, Teresa Blankmeyer Burke, Thomas van Boeckel, Tomas Bogardus, Benjamin Bois, Renee Bolinger, Michael Brent, Liam Kofi Bright, Mark Budolfson, Zuzana Burivalova, Ross Cameron, Sam Carter, Kezia Chuaqui, Ryan Cook, Reyes Coll Tellechea, Andrew Cortens, Christopher Cowie, Sam Cowling, Patrick Kelly Decker, Marco Dees, Neil Dewar, Louis deRosset, Tom Donaldson, Jonathan Dorsey, Jeff Engelhardt, Maegan Fairchild, Marcello Fiocco, Graeme Forbes, Corinne Gardner, Marc Gasser-Wingate, Laura Gillespie, Martin Glazier, Jeremy Goodman, Eric Hansen, Jason Hanschmann, Elizabeth Harman, Mark Harris, John Heil, Michael Hicks, Dien Ho, the members of UC Irvine's Hypatia Society, David Ingram, Alex Jackson, Jonathan Jacobs, Lucy Jordan, Li Kang, Katie Keller, Tom Kelly, Shieva Kleinschmidt, Boris Kment, Patrick Krieger, Kevin Kuhl, Dan Lacker, Kevin Lande, Stephanie Leary, Yena Lee, Uri Leibovitz, Adam Lerner, Sarah-Jane Leslie, Wm Levine, Jon Erling Litland, Heidi Lyne, Una MacDowell, Kate Manne, Beau Madison Mount, Barry Maguire, Ned Markosian, Kris McDaniel, Sarah McGrath, Nicole McManus, John McSweeney, Luke McSweeney, Chris Menzel, Carla Merino Rajme, Alexander Nehamas, Alyssa Ney, Daniel Nolan, Jill North, Graham Oddie, Josh O'Rourke, Josh Parsons, Robert Pasnau, Laurie Paul, Zee Perry, Graham Priest, Kristen Primus, Duncan Purves, Joseph Rachiele, Mike Rea, Bradley Rettler, Lindsey Rettler, Anna Christina Soy Ribeiro, Lisa Rivera, Ben Rohrs, Domenica Romagni, Michael Rota, Robert Rupert, Jeffrey Sanford Russell, Noel Saenz, David Sanson, Karl Schafer, Amy Seymour, Simon Shogry, Matthew Noah Smith, Olla Solomyak, Jack Spencer, Joshua Spencer, Jonathan Squire, Julia Staffel, Jason Stanley, Joey Stenberg, Meghan Sullivan, Noel Swanson, Trevor Teitel, Jonny Thakkar, Lynne Tirrell, Kelly Trogdon, Dimitris Tsementzis, Chad Vance, Leigh Vicens, Judy Vreeland, Polly Wagner, Meg Wallace, Jennifer Wang, Justin Weinberg, Robbie Williams, Alastair Wilson, Annie Wilton, Jack Woods, Audrey Yap, and Dean Zimmerman. Thanks also to audiences at Boston University, California State University-Los Angeles, Oxford University, Pepperdine University, Princeton University, the Society for Exact Philosophy at the University of Miami, the Canadian Philosophical Association at the University of Ottawa, the Mind and Metaphysics NEH Seminar at Washington University, and the Logic and Metaphysics Workshop at CUNY.

Special thanks to Anna Faiola, Jo Kelly, Jason Robinson, Brigid Sudnick, and Joann Zuczek, who together constitute definitely the heart and the brain, and probably most of the other internal organs, of the Princeton Philosophy Department.

Thanks to all the hard workers at Small World Coffee, Fix Coffee, and AP Cafe, for the fuel and the office space. Thanks to Brooklyn Boulders, LA Boulders/Cliffs of Id, The Spot, and especially to John and Ina and Elite Climbing, the Princeton University Mountaineering Club, and the Gunks, for keeping me sane.

I'm especially grateful to all the babies and children who have brought me so much needed distraction and joy in the past five years, especially to Ansel, Hugh, and Dahlia.

Finally, I owe an obvious and rather large debt to both Kant and Wittgenstein. I'm sure I would have hated them both (and I'm sure they would have hated me right back), but they sure have made an impact on me. Thanks guys!

Abstract

In this dissertation, I argue for a novel view in the metaphysics and epistemology of logic, and then answer some of the questions that the view raises. I suggest, first, that logical realism is true—that there is logical structure in the world. It is this logical structure that makes us correct in using certain logics, but not others, to theorize about the world. I then argue that this logical structure is deeply unfamiliar. It doesn't at all resemble any of our concepts, and it doesn't resemble the structure of our languages. I go on to address some of the worries and questions that this raises. First, I argue that the logical realist should think of logics that are candidates for capturing worldly structure as triadic: as consisting in a syntax, a semantics, and a target that that syntax and semantics are jointly aiming to represent. I give an account of when we can know that syntax and semantics as successfully representing the same target, and what we can learn in such cases. Next, I turn to the question of how, if the view I argue for is correct, we can distinguish between metaphysical theories that realists typically take to be saying different things about the world. I show that there is an account of metaphysical equivalence that is consistent with us not knowing what worldly logical structure (or fundamental reality in general) is like, but which still allows us to distinguish between such theories.

Table of Contents

Acknowledgments (p. iii-iv)

Abstract (p. v)

1. Introduction (p. 1-17)

2. Following Logical Realism Where it Leads (p. 18-58)

3. Radical Representation and Logic as Triadic (p. 59-95)

4. Metaphysical Equivalence, Definitional Extension, and Quantifier Variance (p. 97-125)

5. References (p. 126-131)

1. Introduction

This dissertation consists of three papers that attempt to answer some thorny questions about the relationship between metaphysics, logic, and epistemology. I explore the consequences of (and the difficulty in formulating) *logical realism*, the thesis that there is logical structure *in the world*. What does this mean? It is very hard to say, but I take that to be part of the difficulty of the project, and not an obstacle to entertaining the idea of logical realism. My hope is that what it might mean for there to be logical structure in the world will become a bit more clear in this introduction. But it will inevitably become a bit less clear as I go on, for the central claim I make, in the first paper, is that the kind of logical structure that the realist must posit is deeply *unfamiliar*: none of our linguistic or representational resources accurately reflect it. So we can't say much about what this worldly logical structure is *like*; we can't capture its nature.

Still, more needs to be said for us to even begin to understand *what it would be* for there to be logical structure in the world. One way of getting more of a hold on this is to think of some consequences of there being such structure for our metaphysical and metametaphysical theorizing. And one way of doing *that* is to think about how we ought to think of our *language* as capturing, or failing to capture, worldly logical structure.

In light of that, consider the following example: suppose we have two theories, one that describes the world in terms of 'green' and 'blue' and a second that describes the world in terms of 'grue' and 'bleen'. Many of us think that 'green' and 'blue' better respect or reflect the structure of the world than 'grue' and 'bleen' do. Hence, in my terms we think that the 'green'/'blue' theory is more *metaphysically perspicuous* than the 'grue'/'bleen' theory. One way of thinking about the project under consideration here is that it is, partially, an attempt to answer the question: which *logical* notions are metaphysically perspicuous?

I argue, in the first paper, 'Following Logical Realism Where it Leads', that *none* of our logical notions is metaphysically perspicuous: none of them captures worldly logical structure well. But, again, this is a bit deceptive: there are multiple questions here. One is whether our *language* (or *any possible language*) is capable of perspicuously capturing worldly logical structure. A second is whether there *is* any worldly logical structure. A third is, if there is such structure, *what it is like*. My own view is that the answer to this first question is no—we can't describe worldly logical structure in a perfectly metaphysically perspicuous way. That view, which I argue for in the first paper is premised on the answer to the second question being yes: there is worldly logical structure. And a part of the reason that we can't (at least given our current resources) perspicuously capture worldly logical structure is that we have no idea what the answer to the third question is. I show this by arguing that *even assuming* that classical logic is true, we don't know much about that structure at all. It doesn't correspond to the way our concepts and terms currently carve up logical structure.

I don't *defend* my answer to the second question, that is, the claim that there is worldly logical structure. Important work remains to be done in fleshing out arguments in favor of the view. There are, I think, a number of different ways of both motivating and describing logical realism. I won't motivate the view in any depth here, but it's worth quickly laying out some different conceptions of logic and very briefly sketching what it would be to be a logical realist on each of them.

In order to do so, I will quickly and roughly taxonomize (non-mutually-exclusive, and in many ways overlapping) views about what logic is, or is about. First, there is the view that logic has to do with formal properties of mathematical or artificial languages (e.g. predicate calculus, modal logics, etc.). (Call this the 'mathematical view'.) Second, there is the view that logic has to

do with reasoning: that it is about *valid inferences*, about what the *right way* to reason is, and about what *logical consequence* is. (Call this the 'consequence' view.) Third, there is the view that logic has to do with *logical truths*: that it is the investigation of a special set of truths or facts or (that are perhaps the most *general* truths or facts). (Call this the 'logical truth' view, which is most closely tied to Frege's views.) Fourth, there is the view that logic is about the general structure of our thoughts or judgments (call this the 'general thought' view, which is perhaps Kant's view). And there are many other views one might have.

There are also questions that cut across at least some of this taxonomy (though different answers to these questions might pair more or less naturally with each view of what logic is). Two that might bear on what follows, but which I won't discuss in the papers themselves: whether logic is *normative* (e.g. Harman 1986, Field 2009), and whether logic is *a priori* and *exceptional*, or whether our route of investigation into it is just like our investigation into other sciences (and empirically-informed philosophy). This dissertation is, in part, intended to lend support to the anti-exceptionalism view, that is, that our investigation into logic must be continuous with our scientific investigation. There are some interesting things to say about the anti-exceptionalist project more generally, but I leave those aside here and focus on the more narrow project.¹

What would it be to be a logical realist on each of these views? On the mathematical view, it might amount to thinking that the formal properties of mathematical or artificial languages are *in the world*; that the claims we make about logics are ontologically committing. It might amount to denying that formalism in our scientific theories is *mere* formalism: that this

¹ Anti-exceptionalism has its strongest roots in Quine (1951). Anti-exceptionalists (though there is wide variation in the way in which they are endorsing the thesis) also include, among others, Maddy (2002), Putnam (1969, 1976), Priest (2014), G. Russell (2014) (somewhat following Harman (1986)), and Williamson (2013b, 2015). Exceptionalists include (though again, there is wide variation in exactly what is being endorsed) Boghossian (1997), Dummett (1978), Peacocke (1992), Prawitz (1977), and Wright (2001).

formalism is not just a matter of convention, or convenience, or an accident of representation, but that there is something about the world that it reflects (where we aren't necessarily *quantifying* over that something; more on this in the first paper).

On the consequence view, there are at least three ways to be a logical realist: first, thinking that *logical consequence* itself is a worldly relation; second, thinking that there is something else about the world that makes valid inferences valid; and third, if the consequence view is conceived of as involving *norms of rationality*, thinking that there is something about the very structure of the world that grounds these norms, something that they reduce to, that is over and above atomic thoughts, judgments, or propositions.

More generally, if one thinks that logic is normative in any way (which, Harman (1986) notwithstanding, I suspect is the majority view) there is a weak and a strong form of realism about logic one might endorse. The strong form: the very norms of rationality or reasoning that logic is about are metaphysically robust; they are out there in the world, and are in at least some sense independent of us. The weak form: there is something about the mind independent world that *grounds* those norms of rationality or reasoning, or that they *depend on*, or that they *reduce to*. Now, I suspect that only certain versions of this weak form should count as realism. So, for example, if we thought that these norms were completely reducible to, or fully grounded in, atomic facts, perhaps this wouldn't count as the kind of realism I have in mind. But if we think we need something more than atomic facts to get these norms, and that "something else" seems to be something that might count as logical or quasi-logical (for example, non-atomic facts, relations between atomic facts, rules, logical laws, a consequence relation, etc.), then the weak thing *would* count as logical realism in the sense I am conceiving of it.

On the logical truth view, a logical realist might believe that there was something other

than atomic states of affairs that the special class of logical truths were true *in virtue of*. Or, if we think of the logical truths themselves as facts or states of affairs, a logical realist might believe that they have some constituents or some structure over and above whatever non-logical constituents or structure they have. Or perhaps it is enough, to be a logical realist, to hold that there is something in the *world* (and not just a feature of our language or representational systems) that makes those truths *special*. So, for example, consider the logical truth 'it's not the case that San Francisco is in California and is not in California'. And suppose, for a moment, that we are *truthmaker* theorists. What we say about what the truthmaker for this truth is is going to determine whether we are logical realists. If we think that the fact that San Francisco is in California is the *complete* truthmaker for this truth, then we might not be logical realists. But if we think that there is more to its truthmaker than that—that its truthmaker involves, say, the law of non-contradiction, or involves facts about negation and disjunction—then we are likely logical realists of the kind I have in mind here.

On the general thought view, it's hard to say what would count as a realist view of the kind I have in mind—but I also suspect that the general thought view is fairly unpopular among contemporary philosophers. That being said, one might think of Williamson's (2013) view as something that lies somewhere between a more *worldly* version of the general thought view, and a version of the logical truth view: a logic of domain D is concerned with the most general claims about D; D might just be the world; so a logic might just be concerned with the most general truths or claims about the world that there are. I think Williamson's views count as realist in the sense under discussion here, but it would take far more space than I have here to motivate that claim.

As I said, I am not going to argue for logical realism here. But I will mention some

potential motivation for it, first from philosophy of logic and then from metaphysics. First, many philosophers think there is *one true* logic: that there is a fact of the matter about what the *right* logic is, whether it is e.g. classical, intuitionistic, relevant, dialetheic, etc. in nature. (Often this alone, or something close to this, is referred to as “realism” in the literature, but I mean something stronger by that term.) Some philosophers are also logical pluralists, though even some pluralists often don't believe that *any old* logic is correct. One way into logical realism is to ask, what is it *in virtue of which* this is the case? That is, what makes, say, classical logic *right*?

There are all sorts of answers one could give here without entering the territory of logical realism. But many answers do entail some sort of realist view. For example, consider the logical truth view of logic, and suppose we think classical logic is the one true logic. *Why* are the classical logical truths the logically true ones? What makes a dialetheic claim not just not a logical truth, but not true at all? One way into logical realism is to see that we need something *more* than just the non-logical, atomic truthmakers for the logical truths in order to account for certain facts, for example, that the fact that San Francisco is in California makes true the truth that it's not the case that San Francisco is in California and is not in California, but *fails* to make true some other candidate truth.

And we needn't actually be truthmaker theorists to be gripped by this; other views must answer a similar question, it just will be located in a different part of the theory.

Or suppose one likes the consequence view. If one rejects *both* the weak and the strong versions of realism I laid out above, then, I think, it becomes harder to see why logic matters. And it becomes harder to see what this question of what the *one true* logic is is even about if we don't believe in the existence of the consequence relation itself, or of norms of rationality, or of *something* that anchors there being an objective, independent-of-us answer to this question. Of

course, there is room for resistance here. But I see the position that there *is* an answer to what the *one true* logic is, and yet that logical anti-realism is true, as exhibiting a similar kind of tension to that one might think there is between claiming that there is *one true* moral theory, and yet that moral anti-realism is true. (This is not intended as an argument for logical realism—just an observation that the burden seems to be on the logical anti-realist here to reconcile the persistent assumption that there is one true logic with that anti-realism.)

Second, many views in metaphysics, I think, require us to endorse logical realism if we are to take those views seriously, and many *distinctions* made in metaphysics seem to require us to endorse logical realism. So, to the extent that we are metaphysical realists, and that we aren't deflationists about these distinctions, we will need to endorse logical realism too.

What do I have in mind here? I think there are many places in metaphysics where we *implicitly* endorse distinctions that bottom out in questions that have to do with semantics for logical constants, and that any such distinction is one that can't be made without logical realism being true. For we need some reason to think that whatever our chosen semantics is is somehow grounded in, or related to, *reality*, if we are to maintain that a distinction that bottoms out in semantics for logic is a distinction that maps onto something about reality.

There are also places where this sort of thing is more explicitly discussed (though not, to my knowledge, in the terms I am discussing it in), and it would be key to look at these cases if we were trying to generate a more global argument that making metaphysical distinctions requires logical realism. Let me give a very sketchy example. Dasgupta (2009) argues for a view, *generalism*, on which there are no fundamental individuals, and the world is fundamentally qualitative in nature. Generalism is distinct from *individualism* in that it rejects fundamental individuals; it is distinct from *bundle theories* in that it says that, rather than constructing

individuals out of properties, *facts* are constructed out of properties. Dasgupta proposes that the right language for generalists to speak is a language called *G*, which has logical vocabulary consisting of term-functors and an “identity” predicate. And he *explains* these term-functors by appealing to our ordinary understanding of properties and individuals. But this explanation, he points out, appeals to regular old model theory in a way that seems to suggest that we are, in fact, quantifying over individuals. He suggests that in order to become generalists, all we need to do is replace this metaphorical pseudo-semantic explanation with believing that the term-functors and “identity” predicate are *primitives*.

Two things to note here, though I realize this discussion will be a bit hard to follow if one is not familiar with the argument: first, Dasgupta is *explicitly* locating some of the central differences between individualism and generalism *in the semantics for the generalist's logical constants*. (We can see that because the individualist could also use *G* to express her view—and indeed the way that Dasgupta initially instructs us to understand *G* is by appealing to interpreting it using an individualist semantics, then asking us to drop the individualist commitment by taking the logical terms we were interpreting as primitives.) But second, arguably, the generalist could instead use regular old first-order logic with identity to express his view, and instead ask us to *re-interpret* some of our more familiar logical terms. This would require us re-interpreting the existential quantifier, for example, as not quantifying over a domain of individuals at all. (See also J. Russell (ms), Donaldson (2015) for discussion of what it would mean, and whether it is possible, to do so.)

In either case, what is interesting here is that it seems like we can express either theory in either language; choices of languages here might have to do with how *metaphysically perspicuous* the language is (assuming one or the other theory is true). But that doesn't matter

here: what matters is that there is a way to express both theories in one uninterpreted language (where both logical and non-logical terms are uninterpreted!), and that when we engage in the project of interpreting those languages to distinguish between the two theories, the central *difference* between those interpretations is located in the *logical* terms in those languages. And so, the thought goes, the difference between two radically different metaphysical pictures can be located in there being two distinct candidate semantics for logical constants. But then how can we support there being a genuine metaphysical distinction between generalism and individualism without thinking that one of these semantics gets things *right*—captures the way the world really is—and the other gets things wrong? And if we think that one gets things right and the other gets things wrong, this seems to mean that there is something *about the world* that makes one semantics right and the other wrong, and I think that is enough to generate the kind of logical realism under discussion in this dissertation.

Again, (lots) more work would have to be done to develop any argument for logical realism. I just hope to have briefly sketched the beginnings of a few routes we might go down to do so. And again, I simply assume, in the rest of my dissertation, that logical realism is true.

On this assumption, the goal of the first paper is to establish that the logical structure of the world is deeply unfamiliar to us. In rough outline, I ask the reader to consider a dispute between someone who thinks that ' $\vee$ ' and ' $\neg$ ' best respect the structure of the world and someone who thinks that '&' and ' $\sim$ ' do. I argue that it would be objectionably arbitrary to take one or the other collection to be perfectly perspicuous. But taking both to be perspicuous involves positing an objectionable redundancy. This leaves us with two views. *Privileged* says that one collection of our logical constants is perspicuous, but that we can't know which. I argue for the stronger view, *Unfamiliar*, which says that, while there is fundamental logical structure, *none* of our

logical constants respect it—the logical structure of the world is deeply unfamiliar. Consider Benacerraf's argument against a set-theoretic reduction of the natural numbers: there are multiple sets that are equally good candidates for being the number seven; we have no way to choose between them; and so we should reject the claim that numbers are sets, and look elsewhere for an account of the natural numbers. *Unfamiliar* is similarly motivated: there are many equally good candidates for being the perspicuous logical constants; we've got no way to choose between them; and so we should say that none of our logical concepts respect the logical structure of the world. It follows that logical structure is deeply unfamiliar.

This leaves us with a number of very important philosophical questions. I begin to answer two clusters of these, all surrounding issues of representation, in the second paper, 'Radical Representation and Logic as Triadic'. The first cluster concerns how we should think about a logic that is a candidate for being the “one true” logic, e.g. first-order predicate logic, if we are logical realists.²

I argue that the logical realist should endorse a *triadic* view of logic: logic consists in two *representational systems*, syntax (proof theory) and semantics, and a *target*, namely worldly logical structure itself. There is significant precedent for thinking of logic triadically; there is even precedent for thinking of logic triadically where a logic's target is a part of the world itself, rather than being conceptual in nature. So, for example, quantum logic might have (at least once) been thought of in this way: logic is, in some sense, empirical, aiming to capture what the physical world is really like, and, so the argument went, quantum logic does a better job of this

2 Note that, of course, the logical realist needn't deny that there are multiple useful logics that apply in various domains and for various purposes. It is just that there will be a single logic that is the right one for capturing worldly logical structure, or, perhaps, a subset of logics will *bottom out* in the same worldly logical structure, and so perhaps the realist can endorse a kind of pluralism. I suspect, though, that that pluralism will be limited in scope. It can't be that *anything goes* when it comes to which logics correctly capture worldly logical structure. Throughout the second paper, I focus on how we should think about logics that are candidates for being the “one true” logic, and not domain-specific or use-specific logics.

than alternatives. Arguably deontic logic has a *worldly* target as well (if we are realists about the concepts that deontic logic is designed to capture).

The second cluster of questions focuses on representation: how should we think about the relationship between our representational systems (syntax and semantics) of worldly logical structure and that structure itself? And how should we think about the relationship *between* syntax and semantics when we conceive of them as representational? I present a general framework for thinking about the relationship among different representations of a *target* (phenomenon we are aiming at representing). The idea is that we can demonstrate important structural relationships between representations even when we don't know anything about the nature of the target of those representations; and by doing so we can show that two distinct representations successfully represent the same target. (So, this framework should more generally apply to abstract objects and other kinds of entities that we seem to lack empirical access to.) I go on to apply this general framework to logic, showing that one way of thinking about *soundness* and *completeness* is that they demonstrate just such an important structural relationship between representations: they *establish* that syntax and semantics successfully represent the same target.

Of course, this doesn't directly help dissolve the mystery of what that target *is like*, nor does it tell us how to evaluate whether a given syntax and semantics *pair* is the right one—it doesn't tell us how to choose a logic. I suspect that in the end the picture of representation I present in the second paper *does* help us learn something about the nature of worldly logical structure—that it *neither* resembles the syntax nor the semantics of the logic that correctly captures it. But arguing for this claim is a project for further work. I do, however, lay out the options of what we might think this picture of representation does (or does not) teach us about

the targets of our representations.

The third paper, 'Metaphysical Equivalence, Definitional Extension, and Quantifier Variance', addresses a very different question. One worry that arises in the face of any metaphysical picture which involves a kind of division between *the way the world is* and *what we can know about it* is how, and whether, we can know when distinctions we draw in theorizing about the world track genuine features of the world. And the thesis I defend in the first paper seems to create a special version of this problem: we often treat logic as giving us “first principles” which constrain metaphysical reasoning; we also often distinguish between views, in metaphysics and science, by translating loose conceptual distinctions into more precise distinctions that we can characterize using a logical language. The metaphysical meaningfulness of such distinctions is then obscured if there *is* worldly logical structure but we don't know what it is like, and it becomes very hard to regiment distinctions into logical languages if we don't know what we are metaphysically adding or taking away when we do so. It becomes very hard to answer questions like: how much should we *read off* our scientific theories about ontology? Which parts of our (philosophical and scientific) theories are genuinely representational, and which parts are artifacts of representation?

These questions are hard to answer in general, and the third paper can be thought of as standing on its own as an attempt to begin to answer some of them. But it plays an important role in my project as well. Earlier, I suggested that one kind of argument for logical realism is conditional, and takes as its starting point the thesis that metaphysical or scientific realism is true. If such an argument works, metaphysical and scientific realists will need to commit to logical realism in order to support particular distinctions between different philosophical theories, and between different interpretations of physical theories. But I argue, in the first paper,

that logical realism leads to the claim that we don't, and perhaps can't, really know what logical structure is *like*. If that is right, then it is unclear that the logical realist can support these distinctions either. If the logical realist must replace an inability to make metaphysical distinctions with an inability to make *epistemic* distinctions, we might worry that we haven't made any progress; and indeed, some might think that the right conclusion to draw is a deep skepticism about logical realism.

To be clear, I do think that the view I present in this dissertation is inevitably going to have, as a consequence, the claim that sometimes there is a gap between *what we can know about* (or at least: *what we can talk about*) and *the way the world is*, and, perhaps unlike many philosophers, I see no obvious reason to think that that is, globally speaking, a bad-making feature of the view, or a reason to endorse skepticism. But still, much more needs to be said about how we should go about doing philosophy, and particularly metaphysics, in light of these claims.

My third paper can be seen as an exercise in doing so: in it, I claim that there are significant constraints on what we can know and rationally reason about—so, for example, I suspect that one thing that we *cannot* do is provide a single, perfectly metaphysically perspicuous description of fundamental reality, both because we can't describe reality without accidental features of representation creeping in, and because we, I think, can't (or at least deeply *don't*) know exactly what fundamental reality is like—and take that as a starting point for developing a criterion for when we are, and are not, justified in thinking that two purportedly distinct theories *say the same thing about the world*. I take it that the argument I make in the first paper helps support both the claim that we can't describe reality without accidental features of representation creeping in, and the claim that we can't seem to know exactly what fundamental

reality is like. And so it helps set the stage for ruling out accounts of equivalence that depend on the rejection of these two claims (at least, ruling out such accounts as *epistemically* or *methodologically* useful).³

One central question in metametaphysics is whether two purportedly distinct metaphysical theories say genuinely distinct things about the world. In the paper, I offer and consider a necessary condition on equivalence. This condition rules out a thesis, *quantifier variance*, which some think is necessary for certain disputes to be deflated (for example, the dispute between the nihilist and the universalist about composition). There are two very different, but both interesting, upshots of the paper. The first is that the condition I propose gives us a way to block quantifier variance without appealing to naturalness, fundamentality, grounding, or any other metaphysical notions that philosophers have recently worried are metaphysically “spooky” or “esoteric” (e.g. Hofweber 2009). But the second—the one that I am more interested in—is that it, hopefully, motivates a kind of burden-shifting: those who claim that two theories that purportedly say distinct things about the world in fact are equivalent need to give us some sort of *framework* from which we can *see* those two theories as equivalent. The condition I propose is appealing because it allows us to conceive of two equivalent theories as both *contained* within a larger theory that doesn't say anything over and above either one of those theories; but in the end I want to remain open to there being other ways to develop a kind of framework or “perspective” from which two equivalent theories look identical.

I want to end by saying something about the (very) big picture. Though this dissertation

3 The methodology might seem a bit pragmatist. I think it is still far from that, but it also comes apart from a standard way that metaphysicians tend to approach philosophy. The idea is that, if we are doing metametaphysics (and I take the third paper to be a paper squarely in metametaphysics), what we should be aiming for is work that will genuinely and practically guide us in answering the kinds of questions we care about. This doesn't commit me to the view that all our philosophy should be practically useful; but I think metametaphysics ought to be.

concerns the metaphysics of logic, the central, overarching philosophical questions that guide my work might properly be a part of epistemology and, perhaps, the philosophy of language and mind: what are the bounds of *what we can know about the world*? Where does our ability to theorize about the world break down? What kinds of gaps are there between the way we can *talk about the world* and the way it is? How can we (if we can) justify *metaphysical realism*?

How does taking these questions as primary affect what follows? First, even the best candidate, of the three papers contained here, for being about first-order metaphysics, 'Following Logical Realism Where it Leads', is *really* a paper that is mostly about how our concepts, and our language, map onto the world. And there is a reason for that: when I try to examine first-order metaphysical debates, I often find myself thinking that much of what goes on in such debates is a kind of attempt to disentangle (or, alternatively, to reinforce the relationship between!) our language and concepts on the one hand, and the way the world *actually is* on the other hand. And this seems so even when debates are cast as being exclusively about the latter thing. But this is not because metaphysicians are being disingenuous; it is because it seems extraordinarily difficult to reason about first-order metaphysics in a way that manages to consistently respect the boundary between accidents of our concepts and language and *the way the world is*. (And, I should note, this seems to me a global issue in philosophy, but it is perhaps most obviously problematic in metaphysics and metaphysically-oriented philosophy of science.)

So a very first step in my own thinking about any given debate is often to try to figure out where there is good reason to believe that there is mismatch between our language and concepts on the one hand, and the world on the other. I take the first paper to be engaging in that first step. Of course, it also comes to a very substantive conclusion: if there is logical structure in the world, it doesn't look anything like any of our concepts, and it is not carved up in the way that

our language carves it up. I see that conclusion as merely *setting the stage* for thinking about the first-order metaphysics of logic. But does my conclusion make saying anything metaphysical about logic impossible right from the start?

I don't think so, and someday I hope to have something concrete to say about exactly why we can still do metaphysics in the face of such conclusions. But it is the sort of question that is too hard to answer except by slowly dancing around, with the hope of slowly chipping away towards its center. The second paper, 'Radical Representation and Logic as Triadic', attempts to chip away by considering what we can know about our *representations* of some part of the world, when that part of the world is wholly mysterious to us except via those very representations. The long-term question is whether we can learn something about the world itself indirectly, by learning about relationships between our representations of the world. I don't answer that question in the paper. The contribution of the paper is that it lays down and motivates a certain kind of framework for how we ought to think about whether it is possible to learn about worldly 'targets' from facts about the relationships between our representations of those targets, and if so how, exactly, that would work. Much further work would have to be done to answer the long-term question.

The third paper, 'Metaphysical Equivalence, Definitional Equivalence, and Quantifier Variance' attempts to chip away in a very different sort of way. It is very much a paper in metametaphysics, and I don't take it to be making any deeply metaphysical claims. Rather, it explores what happens when we try to think about first-order metaphysical debates by ascending to a metalanguage, indeed a whole meta-framework. It is an exercise that grants, or assumes, that we can't escape certain epistemic and linguistic barriers to knowing about the world, and tries instead to "move up a level"--without thinking that doing so breaks those epistemic and

linguistic barriers--and asks whether doing so can help us understand better the relationships between our theories themselves, but also between our theories and the world.

Certain philosophers have deeply influenced my views here, which might even be properly called 'neo-Kantian'. The reason I make little mention of this throughout what follows is that I also know that Kant (in the far away possible world in which he would bother caring about a Princeton philosophy dissertation in 2016) would turn in his grave at the thought of being associated with much of what follows, in which I help myself to many of the tools and much of the methodology of contemporary analytic metaphysics.

2. Following Logical Realism Where it Leads

Many of us think that two theories can be true, and in some sense equivalent, despite one being *better* than another at describing the world. Consider the theory that describes the world in terms of 'grue' and 'bleen' and the theory that uses 'green' and 'blue'. If one of these theories is true, so is the other. But the latter clearly seems to better respect the structure of reality than the former does. Hence, it matters what choices we make about language when stating our theories: which terms we use, and what the structure of our language is. When it comes to evaluating which terms are more or less *metaphysically perspicuous* (structure-respecting), some cases seem easy: 'electron' is clearly more perspicuous than 'turkey-trout'. Other cases are hard: is 'spacetime point' perfectly perspicuous, for example? To think that there are metaphysically perspicuous descriptions of the world is, as I'll use the term, to be a *realist* about the world having *structure*: if the world has a kind of structure, it must make sense to ask which of our descriptions of it respect that structure.

In this paper, I ask which *logical* notions respect worldly structure. Are '&' and '~' perfectly perspicuous? 'v' and '¬'? '↓' (the 'neither/nor' connective)? What about quantifiers? If the world has logical structure—if *logical realism* is true—then these questions must make sense, and must have answers. The logical realist must think that either some or all of our logical terms are perspicuous—that it makes sense to ask these questions, just as it makes sense for the spacetime realist to ask whether 'spacetime point' is perfectly perspicuous.⁴ And one can be a logical realist, in this sense, without thinking that our logical terms *refer* to anything, just as one

⁴ There is a subtle issue here. I call my own view a realist one, but on my view, none of our (actual) logical terms are perspicuous. So perhaps a slightly more careful way of thinking about this is that it makes sense to *evaluate* our terms for perspicuity if we are realists. It may be that we haven't got the right concepts at all, and so perhaps none of our terms come close to approaching perfect perspicuity. The claim is just that if we think there is some structure in the world, we should think there are better and worse ways of describing it, even if (as I will claim it turns out here) all of our current ways of describing it get things equally *wrong*.

can think that the question of whether 'grue' or 'green' is more perspicuous makes sense regardless of whether we think there are any properties that these predicates refer to.

My primary goal in this paper is to argue that, if logical realism is true, then we are deeply ignorant of the logical structure of the world. This is because one of two views is true. The first view, *privileged*, says that some subcollection of our familiar logical constants are perfectly perspicuous, but that we cannot know which. The second, *unfamiliar*, says that while there is logical structure in the world, none of our logical constants or concepts perfectly respects that structure: we can't talk about it in a perspicuous way with terms like '&' and '~'. While I will offer some reasons for thinking that *unfamiliar* is preferable to *privileged*, the main claim I will motivate is simply that one or the other of the two views is true.

I won't argue for logical realism—the view that there is some logical structure in the world, and that hence we can ask questions of metaphysical perspicuity about our logical vocabulary—here. However, I do want to fend off a potentially serious objection. At first glance, it seems that the logical realist must distinguish between two otherwise equivalent theories, T and T', where T employs ' $\forall$ ', '&', and '~', and employs ' $\exists$ ', ' $\vee$ ', and '~'. But something has gone wrong if we are in a position in which we are forced to ask and answer which of these theories is metaphysically perspicuous, indeed if we are forced to think there is any worldly difference between them at all. These two theories seem to most of us to be paradigmatic *mere notational variants*: the differences between them don't reflect any differences between their metaphysical commitments, and so the question of which is more perspicuous seems like a bad one. If logical realism implies that the question is a good one, the objection goes, then so much the worse for logical realism⁵

5 This challenge to logical realism is not original to me. Sider (2011, ch. 10) to his own view, and argues for *egalitarianism*—the view that all of our logical constants are equally and also perfectly “structural” (perspicuous). I don't engage directly with Sider's argument in this paper, but I should point out that, if my

Both *privileged* and *unfamiliar* give us good explanations for why the question is a bad one, even for the logical realist. According to *privileged*, the sense in which it is a bad question is *that we could never answer it*. The proponent of *privileged* thinks that one or the other collection is indeed metaphysically perspicuous, but that we could never know which. So it is a bad question in virtue of it being unanswerable *by us*. The proponent of *unfamiliar* also thinks our question is a bad one. But for her, this is because *none of* ' $\forall$ ', ' $\exists$ ', ' $\sim$ ', ' $\vee$ ', and ' $\&$ ' respect worldly structure. Rather, the logical structure of reality is radically different from the logical notions we have. We can't accurately capture it using any of our familiar constants, and so none of them are metaphysically perspicuous.

While *privileged* and *unfamiliar* have (in addition to many other virtues!) the virtue of maintaining the realist stance while showing why our “bad question” is indeed bad, both views are surprising, and perhaps initially disconcerting. They entail that we are deeply ignorant about what the world is really *like*. And there are further consequences. Perhaps the most important one is that it seems unlikely that we could construct a perfectly perspicuous theory of the world, the structure of which exactly matches the way the world is actually structured. We don't know exactly what our fundamental commitments should be, and we also don't know exactly what we are committing to, fundamentally, when we state a particular theory (at least given our current supply of logical language), since the logical notions in that theory don't wear their fundamental commitments on their sleeves. And, while *unfamiliar* might give us a bit more hope here than *privileged*, I will offer some reasons, at the end of the paper, for thinking that we are unlikely to discover the perspicuous logical notions.

Before I argue for my thesis, I want to say a few words about why this issue matters.

argument works, it entails that Siderian egalitarianism can't be true. Specifically, egalitarianism is ruled out by one of the principles, *weak non-redundancy*, that I defend in section one.

First, many metaphysicians—whether explicitly or not—seem to be engaged in the project of trying to describe the world perfectly metaphysically perspicuously: to zero in on exactly what fundamental reality is like. Second, in the philosophy of science, both general choices about logical frameworks, and more specific choices about logical constants change the structure and language of our theories. And we're faced with a question of how much to *read off* the resulting differences in structure and language. One of the crucial things we need to know (and be careful about) is what, if anything, to read off a theory's logical structure and language. And answering this question turns on whether logical realism is true, and if so, *how* it is true and how much we can know about the logical structure of the world.

More generally, we often assume that the end of inquiry is achievable and attainable. But if one of the ends of inquiry is to determine what reality is fundamentally like—as both scientists and philosophers often seem to assume--then my thesis here suggests that it is unlikely that we can achieve this end. As I'll suggest at the end of the paper, this does not entail that we should give up on the question of determining what reality is like. Rather, we need to sharpen and clarify the boundaries of the question and recognize our limitations in answering it.

The structure of the paper is as follows: I argue for *privileged* and *unfamiliar* by first, in section one of the paper, discussing a case where a similar challenge arises: when we are evaluating which of a pair of *converse asymmetric relations* like *is above* and *is beneath* is fundamental (or, for the nominalist, which of the corresponding relational predicates 'is above' and 'is beneath' is metaphysically perspicuous). I use this case to motivate two methodological principles. The first, *non-arbitrariness*, says roughly that we shouldn't believe a theory, T, if for every reason we have for preferring T to a distinct theory T', there is an exactly parallel reason for preferring T' to T. The second, *weak non-redundancy*, says that there are no unexplained

necessary connections between fundamental facts. In section two of the paper, I show that logical realism, together with these two principles, entails that either *privileged* or *unfamiliar* must be true. Finally, in section three of the paper, I conclude by discussing the central upshots of the two views, as well as some broader implications.

1. Asymmetric Relations, Fundamentality, and Non-Redundancy

Suppose we are considering whether either of two converse asymmetric relations (such as *is beneath* and *is on top of*), *R* and *Q*, is fundamental. There are various possibilities.

- A) Both *R* and *Q* are fundamental.
- B) *R* is fundamental and *Q* is not.
- C) *Q* is fundamental and *R* is not.
- D) Either *Q* is fundamental or *R* is fundamental, but I can't know which.
- E) Neither *Q* nor *R* is fundamental.

In section 1.1, I argue that neither **B** nor **C** is an option, by motivating a principle I call *non-arbitrariness*: if there are no reasons to think that some theory *T* is fundamental and *T'* is not (and vice-versa), and we have no reason to think that *T* and *T'* are fundamentally equivalent, then we should believe neither that *T* is fundamental nor that *T'* is. In section 1.2, I argue that **A** is not an option, by motivating a principle I call *weak non-redundancy*, which says that there are no unexplained necessary connections between fundamental facts. I conclude that we must believe either **D** or **E**.⁶

⁶ It's important to highlight that we are talking about *asymmetric converse relations* here. You might think I've left a plausible view out: that '*R*' and '*Q*'--the predicates that purportedly name those asymmetric converse relations—in fact are just two different ways to refer to a single relation. I'm sympathetic to this view. But I haven't ignored it. Rather, I've *stipulated* that '*R*' refers to an asymmetric relation like *is above* and '*Q*' refers to an asymmetric relation like *is beneath*. Given this stipulation, there are no possible views other than those outlined above. Suppose, though, that we think there is really only one fundamental relation. Then either the single underlying relation in question is asymmetric or it is symmetric. If it is symmetric, then neither of the relations under discussion *here*—the asymmetric ones—are fundamental. So **E** is true. (Indeed, the best way to flesh out **E**

In section 1.3 I discuss how these options would have to be reformulated for the nominalist and argue that the nominalist must also believe (her analogue) of either **D** or **E**.⁷

1.1 Non-Arbitrariness

I will argue that we should not believe that only *R* is fundamental (**B**) and we should not believe that only *Q* is fundamental (**C**). This is, of course, an epistemic claim, and is distinct from the claim that neither **B** nor **C** is true. It is important to disentangle these two claims, even if in the end one thinks that we can move freely from the epistemic claim to the metaphysical claim that neither **B** nor **C** is true. This is because it is consistent with the claim that we ought not believe **B** or believe **C** that it is permissible to believe the disjunction of **B** or **C**, which is just the claim that one of our relations is fundamental—and, adding to this that we don't know *which* is fundamental just gets us **D**. In the final section of the paper, I will say a bit more about the move from rejecting **B** and **C** to rejecting **D**. Here, I simply want to motivate an epistemic principle that says that we should not believe something when that belief is *arbitrary*—when there is equally good reason to believe an incompatible alternative. This claim—sharpened correctly—seems to me to be very close to analytic. But it is still worth discussing, and, in particular, it is worth making clear what it means for there to be equally good reason to believe an incompatible alternative.

Consider some standard cases of converse asymmetric relations: *on top of* and *beneath*,

is to claim that there is a single underlying asymmetric relation, and that *R* and *Q* are derivative of it.) If the underlying relation is *asymmetric*, then we will be faced with exactly the same issue that **B** and **C** face us with: how could we possibly know *which* of the two candidate relations it was? So there is no real “extra” option to contend with here.

⁷ While my discussion of relations is inspired by Dorr (2004) and Fine (2000), the argument I give is somewhat different from either of their arguments (though much closer to Dorr's, who appeals to something similar to *weak non-redundancy* to motivate his argument), so I won't spend much time discussing them. Additionally, neither of them discusses a reformulation of their argument for the nominalist, and this is crucial for what is to come next (applying the principles we suss out here to the case of logic). Williamson (1985) also discusses this issue.

loves and *is loved by*, *to the west of* and *to the east of*. The thing to focus on, in the case of relations, is that in typical cases, we really don't seem to have any reasons to think one member of the pair is fundamental for which there don't exist similar reasons in favor of taking the other member as fundamental.⁸ We lack reasons for privileging one relation or another.⁹

Let's focus on the case of *is on top of* and *is beneath*.¹⁰ And consider two theories, T^T and T^B , which are identical except that the former mentions the relation *is on top of* and the latter mentions the relation *is beneath*. Now let me introduce a bit of terminology: Let $\text{Fun}(T)$ mean that T is fundamental. So long as we accept that there is a coherent distinction between the fundamental and the non-fundamental, we must accept that $\text{Fun}(T^T)$ and $\text{Fun}(T^B)$ may not be equivalents even if T^T and T^B are. This is because we might individuate theories in a coarse-grained way, but if so, we must treat “it is fundamental that” as a hyperintensional operator. $\text{Fun}(T^T)$ and $\text{Fun}(T^B)$ say different things about the world, even if T^T and T^B do not.

Are there any possible reasons for favoring $\text{Fun}(T^T)$ over $\text{Fun}(T^B)$? I do not think there are: for each reason we might have for thinking that $\text{Fun}(T^T)$, we have an exactly parallel reason for thinking that $\text{Fun}(T^B)$, and vice-versa.

What kinds of reasons am I talking about? Only ones that might have some bearing on the *metaphysical* status of these relations, and in particular, on their fundamentality. It is important to underscore that the problem that arises for asymmetric relations (as well, as I'll argue in section two, as for logical constants) is not a problem about us lacking empirical evidence favoring one relation or the other, or one constant or another—though this is *part* of why there is a problem. The problem is also not that we can't evaluate the balance of reasons, for

⁸ There may be exceptions to this. Perhaps, if the universe has a beginning but lacks an end, that would give us reason to think that *earlier than* is more privileged than *later than*.

⁹ For further motivation see Dorr (2004).

¹⁰ This case comes from Fine (2000).

example because one theory scores better with respect to one theoretical virtue, say, parsimony, and another scores better with respect to a different theoretical virtue, say, explanatory power, and we don't know how to compare the two.¹¹ There are many areas of scientific and metaphysical inquiry in which we run into this sort of problem.¹² But this is not one of them. All of our competing theories will score equally well with respect to each theoretical virtue, and any argument we could make in favor of one would generate an exactly parallel argument for the others.

It follows that we should not believe $\text{Fun}(T^A)$, and we should not believe $\text{Fun}(T^B)$. What is the principle at play here? I take it it is the following:

Non-arbitrariness: If there are no reasons to favor $\text{Fun}(T)$ over $\text{Fun}(T')$ (and vice-versa), and $\text{Fun}(T)$ and $\text{Fun}(T')$ are incompatible, then we should not believe $\text{Fun}(T)$ and we should not believe $\text{Fun}(T')$.

Note again that this is an epistemic, not a metaphysical, principle. It is about what we are licensed to believe, and not about what is true. And if it applies in the case of relations, it will rule out believing **B** and believing **C**. A stronger, metaphysical principle (one which would rule out **D**) is not necessary to generate the central claim of this paper.

That being said, there are reasons for thinking that there is some kind of defeasible methodological principle that moves many of us to “bridge the gap” between the epistemic and

¹¹ For discussion of this point, see Dasgupta (2015).

¹² For example, Bennett (2009) argues that this comes up in disputes over constitution, Dorr and Rosen (2002) suggest something somewhat similar about composition. Though Korman (2010, p. 125) points out that at least in the composition case, it's not true that we have no reason to think that, e.g., there are snowballs but no snowdiscalls (a snowdiscall is “something made of snow that has any shape between being round and being disc-shaped and which has the following strange persistence conditions: it can survive taking on all and only shapes in that range.”), whereas it does seem to be true that (as I will discuss momentarily) we have no reason to think that the number two is identical to $\{\{\emptyset\}\}$ and not $\{\emptyset, \{\emptyset\}\}$. In the first case we've got significant intuitive support for the claim that there are snowballs but no snowdiscalls, whereas there doesn't seem to be any prima facie intuitive support for one view or the other about which set-theoretic reduction of the numbers is true. The narrow cases I'm focused on here are like the set-theoretic case and not like the composition case in this (and many other) respects.

metaphysical versions of *non-arbitrariness*. First, we might have other commitments that lead us to bridge the gap in specific cases, but not in every case. Consider perhaps the most well-known argument involving a non-arbitrariness principle: Parfit's (1984) discussion of fission cases. Parfit asks us to consider a case where my brain is divided between two bodies, 'Lefty' and 'Righty'. The question, in this case, is what happens to *me*. There are four options: I survive as both Lefty and Righty, I survive as Lefty, I survive as Righty, or I do not survive. We can set aside Parfit's conclusions here. I merely want to point out that the argument against the second or third option is that it can't be that I determinately survive as either Lefty or Righty (but not both). And this is not an epistemic claim: there is nothing *metaphysical* (nothing in the psychological and physical facts) that decides between me surviving as Lefty or surviving as Righty, and that's why it can't be that I determinately survive as either.

On one interpretation of Parfit, he is not appealing to some general metaphysical non-arbitrariness principle. Rather, he is basing the claim that it can't be that I determinately survive as Lefty on the claim that *there is no feature of reality* that could make it the case that I determinately survive as Lefty (and not Righty). There being a brute fact that I determinately survive as Lefty would count as being such a feature of reality. So it must be ruled out. In other words: the argument is not that *because* it would be arbitrary to believe that I determinately survive as Lefty (and not Righty), it must be the case that I do not determinately survive as Lefty (and not Righty). Rather, there are additional premises: first, it is not the case that it is a brute matter that I survive as Lefty, and second, that there are no other facts (e.g. psychological or physical facts) that could make it the case that I survive as Lefty. The relevant principle regarding metaphysical non-arbitrariness is analytic: if there is nothing in the world that could make it the case that determinately L or determinately R (but not both), then neither

determinately L nor determinately R. It is just that in order to apply this principle we must buy into there being no Cartesian egos, simple souls, or it being a brute fact how persons persist (at least in fission cases).¹³ We might think something similar about relations if we were convinced that it simply couldn't be a brute fact that *is on top of* is fundamental and *is beneath* is derivative.¹⁴ I take no stand on whether we should commit to this. The point is merely that we might think that in *specific* cases where we had reason to think that we weren't dealing with brute facts, we should transition from the epistemic to the metaphysical version of *non-arbitrariness*.

We might also think there is more general motivation to bridge the gap and move from an epistemic to a metaphysical principle. Consider Benacerraf's argument against set-theoretic reductions of the natural numbers, which relies exactly on this sort of transition: there are multiple sets that are equally good candidates for being the number seven, there aren't any reasons to prefer one set-theoretic reduction over another, and so it follows that numbers are not sets. The fact that this argument seems very convincing to many of us, and that people have adopted it to similarly argue against set-theoretic reductions of e.g. pairs (Armstrong 1986, 1989), propositions (Jubien 2001), and counterparts (Merricks 2003), suggests that many of us at least implicitly endorse the move from the epistemic to the metaphysical claim. The question of

13 It's a messy issue whether this is the correct interpretation of Parfit. One might instead think: the analytic principle is restricted to non-brute, grounded facts. But if that's right, then Parfit's argument is going to fail if personal identity facts are the kinds of facts that could be brute.

14 Fine (2000) also makes this sort of transition. Fine's argument runs through the claim that there is only one *state of affairs* that the sentences 'a is on top of b' and 'b is beneath a' picks out. If that's right, then there has to be a fact of the matter about what the relation is that is a constituent in that state of affairs. He hence rules out the redundancy option. But it's unclear exactly how he's ruling out the arbitrariness option—that it just happens that 'a is on top of b' is the perspicuous description of this state of affairs, because *is on top of* is, in fact, the relation that is a constituent in it. As far as I can tell, we don't get an explicit *argument* from the claim that we ought not believe that *is on top of* is the constituent relation or that *is beneath* is the constituent relation to the claim that neither actually is. But perhaps the idea is that it simply couldn't be a brute fact that *is on top of* was the constituent in this state of affairs, and that there is no possible *reason* that it rather than *is beneath* is.

whether we are justified in doing so in general is one I won't address here, though once I turn to logic I will offer a reason for thinking we are so justified in that case. I merely want to point out that we are often moved by arbitrariness worries to move directly to a metaphysical version of *non-arbitrariness*. we only need the epistemic version to get the arguments in this paper off the ground.

1.2 Non-Redundancy

I have argued against the idea that we should believe that either one, but not both, of a pair of converse asymmetric relations is fundamental. That is, I have argued against options **B** and **C** from section 1. It remains to be shown that both relations cannot be fundamental, that is, that **A** is not an option.

Many people have assumed, usually without argument, that there is some kind of *non-redundancy* or *minimality* constraint on the fundamental. This includes and perhaps begins (via Hume) with Lewis, who says of the perfectly natural properties that “there are only just enough of them to characterise things completely and without redundancy” (1986 p. 60.). On its face, Lewis' claim automatically rules out **A**. If one is already convinced that there is no redundancy at the fundamental level, one should already be convinced that it can't be that both *R* and *Q* are fundamental, because this would entail that there was overlap in the fundamental furniture of the world—that we had *more than we needed* to recover the rest of the world. Why have people been so quick to assume this principle? Perhaps it is motivated by Occamism. Perhaps it is nearly analytic on a certain conception of fundamentality. Claims about the fundamental such as that it is *what God would have to make to get everything else for free* suggest a conception of fundamentality on which it simply couldn't be that *R* and *Q* are both fundamental. For surely

God would need make at most one of R and Q !

Even if we endorse one of these motivations for nonredundancy, it's hard to state the principle clearly: what does it mean for there to be redundancy at the fundamental level? One thing that it is sometimes taken to mean is that the fundamental must be a *minimal supervenience base*. Intuitively: that the fundamental properties, e.g., are those contained in “the” smallest set that everything supervenes on. However, as Sider (1996b) points out, there is no single smallest set that everything supervenes on, which makes it difficult to appeal to this kind of definition of nonredundancy—which set do we choose? Arbitrariness rears its head again.

If we set aside such worries, this strong minimality constraint rules out, as Sider (1996b) points out, that both members of any pair of converse asymmetric relations could be fundamental. And so if such a minimality constraint were correct, we could immediately conclude that it was not the case that both R and Q were fundamental, eliminating the possibility that **A** is true. But it's not clear that we should accept this very strong constraint on either the natural properties or the fundamental, and so not clear that we can rule out **A** so quickly. Sider (1996b, 2011), Eddon (2013), Cowling (2013), and Wang (ms) argue against minimality constraints.¹⁵ And one of Sider's complaints about the constraint is precisely that it rules out what he takes to be plausible cases of fundamental redundancy: logical constants and converse asymmetric relations! So we can't assume a constraint that automatically rules out these cases.

Schaffer (2010) gives an argument for a condition on the fundamental *material objects* that he calls 'No Overlap'. The rough idea is simple. We appeal to a Humean recombination

¹⁵ Cowling (2013) really argues *for* a particular conception of ideological parsimony, but it is one which he suggests can help us avoid arbitrariness worries by allowing us to embrace a multiplicity of ideology when such ideology is interdefinable, and hence by allowing for ideological redundancy in precisely the kinds of cases we're concerned with here. I don't take up Cowling's proposal here since my argument for non-redundancy doesn't run via an appeal to parsimony, but rather via a demand for metaphysical explanation.

principle: the fundamental entities should be freely recombinable. The fundamental entities should be *independent* of one another in that any way of combining them amounts to a genuine metaphysical possibility. But as Schaffer points out, there are modal constraints on overlapping entities: “Consider two overlapping homogenously red circles, each of which individually could have been green.” (2010, p. 40) We can see the problem: circle A can't retain its parts while being green if circle B retains its parts while remaining red (this would entail that the overlapping bits of A and B are both homogenously red and homogenously green). So we can't both endorse free recombination and allow for overlapping fundamental ontology.

Can a similar argument be made with respect to relations, and in particular, pairs of converse asymmetric relations like *R* and *Q*? Recombination is typically characterized as a principle about regions of space or spacetime (e.g. Lewis 1986 p. 70-90) and how the entities that live in spacetime can be recombined, whether they be events, individuals, states of affairs, and so on. But we need something much more general, along two different dimensions, here: first, we want our principle to apply to different *kinds* of entities, and second, we don't want it to discriminate between nominalist views and realist ones. And so it is perhaps misleading to characterize the search as one for a *non-redundancy* principle. What we really need is a principle requiring necessary connections between *any* kind of fundamentalia to have metaphysical explanations. Here, I can't possibly motivate such a general principle. But I will try to motivate it for facts, staying neutral on what kind of entity a fact is (state of affairs, true sentence, proposition, etc.), except to claim that fundamental facts only contain fundamental entities (be they linguistic, concrete, abstract, etc.).

The principle I propose is the following:

Weak Non-redundancy: There are no unexplained necessary connections between fundamental facts.

Why is this constraint called 'non-redundancy' when it seems to be about unexplained necessary connections, and not redundancy? It is clear from Schaffer's argument that it is *overlapping* fundamental ontology that will create these unexplained necessary connections. I will show that there is an argument against redundancy that extends first to states of affairs, and second, not just to fundamental ontology but also to fundamental ideology—to the nominalist's fundamental language. Why is *weak non-redundancy* 'weak'? Because it's far weaker than a minimal supervenience base constraint: it allows, in principle, for there to be necessary connections between fundamental facts—it just requires that those connections themselves have explanations.

Motivating *weak non-redundancy* is straightforward if we reify relations. I will first outline this argument, and then, in section 1.3, will show that the nominalist is in no better of a position to reject non-redundancy than those who reify relations are.

First let's get clear about what denying *weak non-redundancy* in the case of asymmetric converse relations amounts to. Suppose that we have an ontology of states of affairs, and suppose that R and Q are a pair of asymmetric converse relations, and are both fundamental. Then, for every fundamental state of affairs (e.g. Rab) that involves the instantiation of one of our relations, there will be a corresponding fundamental state of affairs that involves the instantiation of the other (Qba). And they will necessarily co-obtain. It is implausible that there is *no* explanation for this necessary connection. On such a view there is simply no reason at all why Rab and Qba travel together, and nothing to explain why R and Q are so intimately related. They are just two relations that happen to always travel together through modal space.

But the fact that most of us will not find the “no explanation” view plausible does not

mean that all of us will be moved by Humean Recombination—that the necessary connection simply can't obtain because the fundamental facts must be freely recombinable. So the question, in the case of relations, is whether there are any plausible candidate explanations for the fact that Rab and Qba necessarily co-obtain. I will consider a few such explanations here.

First, we might claim that Rab and Qba are identical—that they are the same state of affairs, described two different ways. This seems hard—perhaps impossible—to maintain without collapsing into the position that there is a single underlying symmetric relation that is the real fundamental relation instantiated in that state of affairs—that is, without collapsing into endorsing option E. It is of course open to us to say that 'Rab' and 'Qba' pick out the same underlying state of affairs (indeed, this is what Fine (2000) claims). But when we ask what *relation* is a constituent in that state of affairs, I don't think we can consistently claim that there are *two distinct* asymmetric converse relations that are both constituents of it. For if we could, why couldn't we genuinely (rather than merely notationally) separate out two different states of affairs here? Surely if being a renate and being a cordate are distinct properties, the state of affairs of Izzy being a renate is separable from the state of affairs of Izzy being a cordate on exactly these kinds of grounds.

Instead, this sort of view pushes us into thinking that what's really going on is that we've got two distinct relational *predicates*, 'R' and 'Q', which both refer to the same underlying (presumably symmetric) relation. One option is to deny the 'presumably symmetric' claim, but then we would be back to grappling with how to hold that 'R' and 'Q' both referred to the same *asymmetric* relation in the face of *non-arbitrariness*.¹⁶ It's important to keep track of the dialectic

¹⁶ One might instead deny that states of affairs have relations as constituents, for example, by thinking that states of affairs are 'chunks' of reality that make sentences and propositions true. But on such a view, it's hard to see why we would need to posit fundamental relations in the first place, if we thought states of affairs were fundamental.

here: to admit that there is a single underlying symmetric relation and a single state of affairs is to reject that both R and Q are fundamental, or that either is (remember that we are *assuming* that they are asymmetric relations), so it is already to rule out *all of* our options **A**, **B**, and **C**.¹⁷

However, as I'll discuss in section 1.2.1, analogues of the kinds of moves outlined in the last two paragraphs are available to the nominalist, and it is a more delicate matter to reject them there.

There might be other options available. Perhaps it is a *metaphysical law* that R and Q always travel together through modal space. Two things work against this claim. First, there's an important sense in which metaphysical laws *seem contingent*. Consider standard examples of metaphysical laws: laws of mereology, the axioms of set theory, perhaps direction-of-ground. We can conceive of, understand, and imagine situations in which those laws fail to hold. We just don't count them as metaphysically possible. But it is inconceivable that R_{ab} and Q_{ba} don't necessarily co-obtain. Thus what we are trying to explain here seems importantly different from cases in which we normally appeal to metaphysical laws.

Second, metaphysical laws don't generally give us *symmetric* explanations. They are usually the kind of thing that serve to underpin or explain either the nonfundamental in terms of the fundamental, or to underpin or explain cross-temporal relations (cause and effect). We might posit laws of mereology that govern relations between larger objects and the smaller objects that compose them, but assuming we are already in the fundamentality mindset, we will think that these kinds of laws relate the fundamental to the non-fundamental. So here again there is a difference between our standard use of metaphysical laws and what we need them to do. For these reasons, I don't think metaphysical laws can do the work required here.¹⁸

¹⁷ Fine (2000) uses this claim to argue for the neutral relation being the real one.

¹⁸ One might also worry that if one had a Humean conception of laws, this strategy would result in a circular explanation. I don't have the space to argue for this claim, and there may be ways to reject it. For a somewhat parallel debate in the philosophy of science/physics, see e.g. Loewer (2012), Miller (2015), Hicks and van Elswyk (2015), Lange (2013), Bhogal and Perry (2015), though many of the solutions explored there won't

One might instead want to claim that it is part of the *essence* of *R* that wherever it is instantiated, so is *Q*, and vice-versa.¹⁹ The first thing to note here is that this kind of explanation requires commitment to a non-modal account of essence. If essences bottom out in modal claims, then we can't use them to explain modal claims! So we need to take something like Fine's approach, on which essence facts *metaphysically explain* modal facts, and not the other way around. At first glance, so long as one is already on board with this kind of Finean approach, this seems like better approach than appealing to metaphysical laws—in part, perhaps, because essence facts don't seem contingent in the same way that I just suggested laws do. But the essence approach doesn't fare any better.

There is something fishy about positing symmetrical (“reciprocal”, in Fine's terms) essences in this way, given that we are *using* the essence facts to do explanatory work. To see why, we first need to get on the table Fine's distinction between *constitutive* and *consequential* essence. Fine says: “An essential property of an object is a constitutive part of the essence of that object if it is not had in virtue of being a consequence of some more basic essential properties of the object; and otherwise it is a consequential part of the essence.” (p. 57.) Here, Fine is very clear that by 'consequence' he means *logical* consequence, in a rather narrow sense.

Fine himself says, about his concept of constitutive essence: “It seems very reasonable that reciprocal essences should be forbidden under the constitutive conception; for how else could a noncircular account of the essence of things be given?” (1994b, p. 65.) The idea is that we don't want a circular account of any individual object's essence, one which builds in the identity of the object in question, since essences are doing our basic explanatory work. But we will run into this problem if we allow reciprocal *constitutive* essences.

work if we are focused entirely on a metaphysical, and not a scientific, case.

19 Feel free to translate this sentence into your favorite way of stating essence facts (so long as it doesn't run into the circularity problem that modal conceptions of essence do).

And while Fine claims that one needn't think that this is also true of consequential essence, it is very hard to see how the details would work here. First, it's unclear how we could get from consequential to constitutive essence if we took the former, rather than the latter, as basic, whereas it's easy to see how we could go the other way around (as Fine himself notes). But, then, it's unclear how R and Q could possibly end up involved in each other's respective consequential essences, given that they are not strictly *logically* related to one another at all. It's not a *logical consequence* of anything about the nature of R that whenever Rab obtains, so does Qba . But if that's right, and it is constitutive essence that is basic, then there is no way for Q -facts to end up in the consequential essence of R , and vice-versa, without already being in one another's constitutive essences.²⁰

If, instead, we want to claim that Q is just 'logically constructed' out of R , then, I think, we're back to claiming that one, but not both, of R and Q is more fundamental, since we are claiming that Q just arises out of the consequential essence of R . And it's easy to see why we can't say that Q is 'logically constructed' out of R and vice-versa: we'd end up with two relations that are both mere logical constructs, with nothing upon which they ontologically depend. It proves to be very difficult to invoke essences as doing genuine explanatory work in the case of relations, and these issues will extend to the case of the logical constants.²¹

20 Rosen (ms, p. 9-11) argues for a 'consequential essence first' approach which may well avoid this worry. But such an approach comes with an epistemic cost that seems to me too great—we have great difficulty isolating an entity's constitutive essence if we start out with its consequential essence. (This epistemic worry also appears to be at least partly what motivates Fine to give a 'constitutive essence first' account.) But that is an issue for another day. I suspect that Rosen can probably account for the issue under discussion in a way that Fine cannot, that this is a significant advantage of Rosen's account, and the details would have to be hashed out, which I don't have space to do here.

21 We might additionally have a related worry: when we appeal to the claim that it is part of the essence of scarlet that anything that is colored scarlet is colored red, we are appealing to an asymmetric essence fact (it is not the case that anything that is colored red is colored scarlet), in order to explain an asymmetry between determinates and determinables. When we claim that it is part of the essence of Socrates' singleton set that it contain Socrates, but not part of the essence of Socrates that he be the sole member of his singleton set, we are appealing to an asymmetric essence fact in order to explain an asymmetry between a more fundamental and a less fundamental entity. These kinds of essence facts are posited because they play an explanatory role that, given certain commitments, we cannot do without filling. But it is not clear that we can't do without filling that role when it

One alternative is to take the necessary connection to be grounded in something like the plurality of the natures of R and Q (of R and Q 'taken together'), but not in either the nature of R nor the nature of Q . However, if we accept a fairly natural connection between fundamentality and ground, we should reject that it could be the case both that this was true, and that R and Q were individually more fundamental than R and Q taken together. For accepting that both these claims could be true means that the grounding chain for this necessary truth would *bottom out* in a non-fundamental entity (the plurality of R and Q). And we might think that grounding chains either must bottom out at the fundamental level, or, if we deny that there is such a level, continue forever. Denying this connection between fundamentality and ground conflicts with many conceptions of fundamentality, e.g. that we can 'get everything else for free' from the fundamental.

I have argued that *weak non-redundancy* motivates a rejection of **A**: that we cannot think that both of a pair of converse asymmetric relations are fundamental. And I have argued that *non-arbitrariness* motivates rejecting **B** and **C**. But I have assumed that we are talking about our fundamental *ontology*. What happens if, instead, we reject the reification of relations and ask whether the same principle applies to fundamental *ideology*?

1.3 Nominalist Views

Do *non-arbitrariness* and *weak non-redundancy* apply in just the same way if we don't reify relations? We have general reasons to think that, if we make our ideology do the same kind of work for us that ontology does, problems with our ontological commitments will be recreated at the level of ideology: that we will be pushing the bump around under the rug. More specifically,

comes to symmetrical or 'reciprocal' essence claims, and so one might be suspicious that our theoretically motivated tool is overreaching in these cases.

if despite not reifying relations, we require our relational *predicates* to do very similar work to the work that relations would otherwise be doing, we should be suspicious of claims to the effect that the very same worries will not arise for the nominalist that do for the view on which we ontologically commit. Even claims of a nominalist's advantage with respect to parsimony considerations are problematic: if the nominalist wants her theory to do the same work as the realist, her theory is going to end up being *ideologically* unparsimonious in just the same way that the realist's theory is *ontologically* unparsimonious.²²

Of course, this might only be the case if we take our ideology metaphysically seriously. What does this mean? The nominalist who wants to be able to claim that the 'green'/'blue' theory of the world is *more metaphysically perspicuous than* the 'grue'/'bleen' theory of the world must think that, even if there is nothing that these predicates *refer* to, some of them better reflect the way the world is structured than others do. She just doesn't want to *reify* this worldly “structure”. Any picture on which we can be nominalists while taking our ideology seriously will suffer from many of the same problems that their reifying counterparts do.²³

To show that this is the case here, first, let's translate our options **A-E** into nominalist-friendly options. I'll do the translation in terms of fundamental *predicates*.

A*) Both 'R' and 'Q' are fundamental.

B*) 'R' is fundamental and 'Q' is not.

C*) 'Q' is fundamental and 'R' is not.

D*) Either 'Q' is fundamental or 'R' is (and not both), but I can't know which.

E*) Neither 'Q' nor 'R' is fundamental.

There is still a pressing question about an unexplained necessity when we move to nominalism:

²² For a discussion of this issue applied to a different debate, see Turner and Dasgupta (forthcoming).

²³ See, e.g., Lewis (1983), Sider (2011), Turner (2016).

why is 'Rab' true in our fundamental language whenever 'Qba' is, and vice-versa? If *non-arbitrariness* applies to reified relations, it also applies to nominalist views that are committed to fundamentality: what reasons could we possibly have for thinking that 'Rab' is a more perspicuous description of reality than 'Qba' is? We can quickly rule out believing **B*** and believing **C***.

What about **A***? **A*** is ambiguous between two readings. It's hard to talk about either of these readings in nominalist-friendly terms, so instead I'll speak as though I'm reifying “structure”, but this will be immediately followed by an argument for the claim that this is necessary to get the commitments of the “fundamentality-friendly” nominalist on the table. On the first reading, **A*** is true because the world has a single kind of structure, and 'Rab' and 'Qba' are equally good, and are both the best, expressions of that structure. But note that this is roughly akin to the platonist saying that 'Rab' and 'Qba' pick out a single state of affairs, that they both equally perspicuously represent it, and that there is no more perspicuous representation of it. So it is most like the platonist's option **E**, and hence supports my overall conclusion. The nominalist just adds an additional claim: that there is no better way to represent the world than with 'Rab' and 'Qba'. I take this view to be friendly to my conclusion here—it is roughly the nominalist's version of the platonist's concession that there is only a single, *symmetric* relation *in the world*. And note that it doesn't fall prey to *weak non-redundancy*: while there are necessary connections between the sentences 'Rab' and 'Qba', there is an explanation for those connections, namely that they are equally best representations of a single kind of underlying structure.

Notice, though, that for the nominalist to claim that 'Rab' and 'Qba' are not just equally fundamental, but are tied for being *the most fundamental* sentences, suggests that she does not think we have the linguistic tools to express something more reflective of the single way that the

world actually is. 'Rab' and 'Qba' can't both be *perfectly perspicuous* ways to describe that underlying reality, for a perfectly perspicuous linguistic representation would not reflect any extra structure that the world does not have, and if 'Rab' and 'Qba' both are equally (and most) fundamental, they do involve extra structure (since they have conflicting sentential structure).²⁴

On the second reading, A^* is true because the world is multiply-structured: it *has* (in some non-reified sense!) Rab structure and Qba structure. But this falls prey to *weak non-redundancy*: if the world has Rab structure and Qba structure, then it has overlapping and redundant structure. Even if we don't want to reify this structure, we can appeal to weak non-redundancy here in virtue of its being neutral with respect to what it ranges over. There are necessary connections between 'Rab' and 'Qba'. And those connections are unexplained, since we can't appeal to the “they most perspicuously describe the same structure” explanation.

At this point my reader may be horrified. What kind of argument am I giving that appeals to “structure” in explanation, quantifies over it, counts it, etc., when it is precisely that structure that we are trying to be nominalists about? I claim: it's a *presupposition* of the fundamentality-friendly nominalist that we are forced to think about things this way, while keeping in mind that we don't *really mean* to be reifying “structures”. So, insofar as my reader is horrified, her horror ought to lie with fundamentality-friendly nominalism generally, and not with the claim that *if* it is true, there are, genuinely, two different ways that the world could be that would make A^* true.²⁵

24 The nominalist may have an available response here that involves denying that these two descriptions have distinct sentential structure. I don't discuss this move here because it (importantly) *can't* be made in the logic case, which is one reason to think that there is an even stronger case to be made there than here. I'll say more about this when I turn to logic.

25 It's quite hard to make the general idea here clear. But there is significant precedent, not just for assuming it works, but for making use of it in conjunction with some or other notion of fundamentality or naturalness. See, for earlier ancestors, e.g., Quine (1948), Armstrong (1978). More directly on this issue, see Lewis (1983), who suggests that naturalness can apply to the nominalist's predicates, Sider (2011), and Turner (2016, section 1.1.3) for an especially clear discussion of this issue.

To clarify, think about an analogy to a slightly more familiar issue in modal metaphysics. Consider two different views about *modal properties*. One view says that modal properties must have their source in categorical features of the actual world: that, in some metaphorical sense, we could just *look at* an object and be able to “read off” its modal properties from its categorical properties: what *makes* an object have the modal properties it does is just *whatever the object is actually like*, what non-modal properties it has. Another view says that modal properties can be *brute*: that objects can have modal properties that aren't grounded in any of their categorical properties.²⁶

There is a less familiar distinction to be made when we are thinking about *representation*, and in particular, what it is to be a perfectly metaphysically perspicuous *representation* of the world. One view says that a perfectly perspicuous representation of reality is perspicuous at least partly *in virtue of* the world actually being a certain way (where we needn't be reifying “ways”, and where we are being very loose about what 'in virtue of' means): having a certain nature or structure. A second view says that *nothing worldly* contributes to which description of the world is perfectly perspicuous. The claim is that we can't hold the latter view if we are taking our ideology metaphysically seriously, e.g., if we are fundamentality-friendly nominalists.

For consider what it would be to endorse it if we were property nominalists: 'green' and 'blue' more perspicuously describe the world than 'grue' and 'bleen' do, but this has nothing to do with how (language-independent) reality *is*. This is in direct conflict with the initial motivation for distinguishing between *metaphysically* more and less fundamental language: we thought we were doing so to reflect something about *the world*, not to reflect purely linguistic facts. Moreover, it makes the fundamentality-friendly nominalist's theses metaphysically uninteresting:

²⁶ e.g. Williamson (2013a).

she thought she was telling us something about metaphysics, about the world, but it turns out that she is just reporting facts about our language, that don't have their source in the language-independent world. So the only way to make the fundamentality-friendly nominalist's thesis metaphysically substantive, and in line with her motivation for adopting it, is to adopt the first view: a perfectly perspicuous representation of reality is perspicuous *in virtue of* the world being a certain way.²⁷

While it's difficult to say what this amounts to in nominalist-friendly language, it's clear that it's a necessary feature of the view. To take *ideology* metaphysically seriously just is to say that the world somehow *metaphysically supports* certain choices and distinctions in language—but, in most cases, it is also to take this relation of metaphysical support as a primitive, which makes the challenge of spelling out just how this works unmeetable.

I have argued for *non-arbitrariness* and *weak non-redundancy*, and have shown that they extend both to realism and nominalism about relations. In what follows, I turn to the case of logical constants, argue that both principles apply there as well, and discuss what we ought to conclude about logic.

2. Logical Constants

What is it to be a realist about logic? On the characterization of logical realism that I am considering here, it is to think that the world has, in some unspecified sense, logical structure,

27 This style of argument is perhaps uncharitable to the neofregean, and in particular is incompatible with Rayo's (2013) views. However, I have the same questions about his view as I do here. Rayo claims that there are certain “just is” statements, such as *for there to be a table there just is for it to tableize*, that are true, and that this view is compatible with the kind of “metaphysicalist” (in his terms) view I am discussing here—one which says that different languages are more and less apt for carving nature at its joints. Elsewhere, I argue that these two claims are jointly incompatible, for very similar (but expanded) reasons to those I give here. In particular, there's an extended argument to be given for the claim that the metaphysicalist ought to buy into the claim that *being more* or *being less* joint-carving, while a feature of languages or linguistic or representational entities, must have its source, in some sense, in the language-independent world, and that this is makes the metaphysicalist's views incompatible with (at least some of) Rayo's views.

and that there are hence metaphysically more and less perspicuous ways to use logical vocabulary to talk about the world. I will show that it turns out that our collection of logical constants is *redundant* in much the same way that R and Q are; and that choosing between subcollections of constants faces us with the same sort of arbitrariness worries that choosing between asymmetric relations does. In the case of logic, we are forced to similar solutions to those that relations face us with: either a privileged subcollection of logical constants is perfectly perspicuous—they get the logical structure of the world just right, but we can't know which (*'privileged'*), or the fundamental logical structure of the world is deeply unfamiliar and doesn't resemble any of our logical constants (*'unfamiliar'*).

Before I argue for *privileged* and *unfamiliar*, I'm going to distinguish between two different kinds of logical realism (which correspond to the two views discussed in the previous section, realism which reifies relations and realism which nominalizes them), which will help clarify what I mean by 'structure'. (Though my conclusion here entails that I can't entirely clarify this notion!) First, one might think that expressions like '&', 'v', '¬', etc. *refer* to individual entities.²⁸ What kind of entities? The most likely candidates are truth functions, or the “worldly” equivalent of truth functions (perhaps: whatever the *logical* constituents of states of affairs are). I'll call this kind of view—roughly a view where commitment to logical constants involves some kind of ontological commitment—*ontological realism*.

One might instead think that while '&', 'v', '¬', etc. are *syncategorematic*—that they don't refer at all—that they nevertheless play an important role in carving up reality, that is that they are bits of ideology in the sense that our non-logical predicates were in section 1.3. On this view, the constants don't just carve reality up in a way that's convenient for our own purposes, or that

²⁸ I am loose-talking here, glossing over some issues with reference. But I don't think it matters for the purposes of the argument.

depends on the way our minds or language are structured. Rather, certain logical constants perfectly represent the way reality is—they are *joint-carving* bits of ideology.²⁹ One might think, for example, that there is conjunctive fundamental structure, that is, that conjunction carves nature at its joints, but that there is no sheffer stroke (the “not/and” connective) structure—the Sheffer stroke does not. I will call this kind of view—the kind of view on which we are ideologically but not ontologically committed when we talk in terms of the constants—*ideological realism*.

Both of these views are forms of logical realism. They might be hard to get a grip on at first, but just compare them to the case of 'green', 'blue', 'bleen', and 'grue'. Ontological realism is like a view which says that at least some of these predicates refer to genuine properties. There is then a question of whether some of these properties are more or less fundamental than others (or alternatively, which predicates genuinely refer). The ontological realist will similarly be faced with a question of whether certain logical entities are more fundamental than others. Ideological realism is like the view discussed in section 1.3: 'green', 'blue', 'bleen' and 'grue' are predicates which don't refer to properties, but yet still do better and worse jobs at carving nature at its joints. And just as the property nominalist can think that 'green' and 'blue' are more *natural* or *fundamental* or *structural* predicates than 'grue' is, so the ideological realist might think that certain logical constants carve the world up more perspicuously than others.³⁰

For now I will take ontological realism to be my target. I will later explain how the problems I raise extend to ideological realism. And I'll assume that classical logic is the “one true logic”. Similar issues arise for most logics, though some do better than others.

Suppose that ontological realism is true. It follows that *being a correct fundamental*

29 For discussion of such a view see Sider (2011), Turner (2016, introduction).

30 For more on this sort of view see Lewis (1986), Sider (2011).

theory is an extremely fine-grained matter. Whether the theory that is formulated using $\exists$, $\sim$, and $\vee$ (call it T) is the correct fundamental theory comes apart from whether the theory that we would naturally call equivalent to T , but which is formulated using $\forall$, $\sim$, and $\&$ (call it T') is the correct fundamental theory. So T and T' are fundamentally distinct.³¹ (I'll continue with the convention of treating 'Fun(T)' as shorthand for ' T is the correct fundamental theory'.) Why does this follow? Fun(T) and Fun(T') have different fundamental ontological commitments. Fun(T) isn't committed to there being any fundamental conjunction at all. And Fun(T') is. And vice-versa for disjunction.³²

If Fun(T) is a distinct claim from Fun(T'), then it makes a *metaphysical* difference, and not just a pragmatic one, which one we choose to accept. Given that, we are faced with a familiar problem: there doesn't seem to be any possible reason we could have for believing Fun(T) rather than Fun(T')--in every case we can generate a parallel reason for believing Fun(T'). And vice-versa. What do we do? Accept Fun(T)? Accept Fun(T')? Neither? Both? As in the relations case, there are five things we might believe:

- A) Both Fun(T) and Fun(T').
- B) Fun(T) and not Fun(T').
- C) Fun(T') and not Fun(T).
- D) Either Fun(T) or Fun(T'), but I can't know which.
- E) Neither Fun(T) nor Fun(T').

I claim that we should only believe **D** or **E**, which are really just the views I introduced earlier as *privileged* and *unfamiliar*. *Privileged* is the view that some subcollection of the logical constants

31 Don't read much into what I mean by 'theory' here. I might as well have said "description in a language", or something like that.

32 One might take this fact—the fact that ontological realism is committed to T and T' saying distinct things about the world—to itself be a reductio of fundamental ontological realism. But that is not what is at issue here.

is perfectly perspicuous, but that we can't know which. And that is just what **D**, suitably generalized, amounts to. *Unfamiliar* is the view that the fundamental logical structure of the world looks nothing like *any* of our logical constants, and so none of our constants perspicuously represent reality. **E**, suitably generalized, together with the claim that logical realism is true, entails *unfamiliar*. In what follows, since I am assuming logical realism, I will use '*privileged*' and '*unfamiliar*' to talk about these last two options.

2.1 Non-arbitrariness and Logical Constants

In this section, I assume ontological realism and argue that we ought not believe **B** and we ought not believe **C**. Do we think that '&' or 'v' refers to a fundamental entity? I claim that we can't generate any reasons to favor one for which we don't have a parallel reason to favor the other, and so *non-arbitrariness* applies.

What would such reasons look like? There are three kinds of considerations I can think of that we might hope would give us non-parallel reasons, and none are successful at doing so. The first are parsimony considerations. Perhaps we should take theories with the fewest number of logical primitives to be more likely to be fundamental. But then we will be stuck trying to compare a theory formulated with the down dagger (the single connective that expresses “neither/nor”) with one formulated with the Sheffer stroke (the single connective that expresses “not/and”). Each is equally parsimonious with respect to logical concepts, and we have no way to arbitrate between the two.

The second is psychological, and concerns what concepts are easiest for us to work with. Perhaps, it is easier for us to reason using the connectives $\sim$ and $\&$ than it is for us to reason using the connectives $\sim$ and v . But it's very hard to see how such pragmatic considerations about

ease of use could generate *metaphysical* reasons for preferring one theory to another. Any argument to that effect would be quickly debunked by considering (actual or possible) reasoners for whom $\sim$ and $\vee$ were easier to reason with than $\sim$ and $\&$.

Finally, and relatedly, we might make an appeal to conjunction's *seeming* like the most natural connective as a reason to claim that, say, conjunction and negation are the fundamental constants. There are two responses to this. First, even if we grant that, for some reason, conjunction and negation seem to be the most natural constants, it is hard to see why we then wouldn't have a problem deciding between conjunction and negation on the one hand and the Sheffer stroke on the other. If anything, it seems that parsimony considerations would push us towards treating the sheffer stroke as fundamental. But then the intuitive motivation for treating conjunction and negation as fundamental is lost. The second response illuminates the first: while it might seem that conjunction is more natural than the other constants, the same thing is simply not true when it comes to negation. The insistence that it is conjunction and negation that are the most fundamental is *really* an intuition that conjunction is the most natural, and then a realization that conjunction alone has very little expressive power, but that adding negation in gives us just as much expressive power as any other collection of (propositional) connectives. So it seems that if one is going to take seriously this sort of intuition, one ought to claim that conjunction is the only fundamental logical constant. But such a view is severely impoverished.³³

There are no reasons to prefer one of T to T' without automatically having a parallel reason for preferring the other. So *non-arbitrariness* applies. We should reject that either Fun(T) or Fun(T'). So, as in the case of relations, neither **B** nor **C** are options. And it is again easy to see how this style of argument would extend to the ideological realist as well as applying to the

³³ Note the similarity between the issues with negation that arise here and those that arise for truthmakers, e.g. in Armstrong (2004).

ontological realist. It remains to be shown that *weak non-redundancy* generalizes to logical constants, which means that we cannot accept **A**: that *both* $\text{Fun}(T)$ and $\text{Fun}(T')$.

2.2 Weak Non-redundancy and Logical Constants

We are considering consequences of holding that logical *realism* is true. This means that our use of the constants reflects something worldly: either they directly refer to entities in the world, or they best capture the way the world is structured. One way this could be true is if we had an ontology of not just atomic but also logically complex states of affairs. That view might be unpalatable, but that doesn't matter here. I'll use it to argue against **A**, and it will be easy to see how to discharge the assumptions I make without sacrificing the claim that **A** can't be right. Let's also take on board the idea that states of affairs are *truthmakers* for true sentences. Now, let's suppose that **A** is true, and ask what the *truthmakers* for a sentence of T , ' $\sim A \vee \sim B$ ', and the corresponding sentence of T' , ' $\sim(A \& B)$ ' are. More specifically, let's ask what the *most fundamental* truthmakers for these sentences are.

Suppose that the state of affairs $\sim A \vee \sim B$ was the most fundamental truthmaker for both sentences. Could **A**, the claim that T and T' were both fundamental, be true? No. ' $\sim A \vee \sim B$ ' and ' $\sim(A \& B)$ ' are clearly not equally perspicuous representations of the state of affairs $\sim A \vee \sim B$. We'd have the same problem if we took $\sim(A \& B)$ to be the fundamental truthmaker.

So what could the fundamental truthmakers for ' $\sim A \vee \sim B$ ' and ' $\sim(A \& B)$ ' be, if **A** were true? There are three remaining options: first, *both* the state of affairs $\sim A \vee \sim B$ and the state of affairs $\sim(A \& B)$ are truthmakers for both sentences, second, *neither* state of affairs is, and third, $\sim A \vee \sim B$ makes ' $\sim A \vee \sim B$ ' true, and $\sim(A \& B)$ makes ' $\sim(A \& B)$ ' true. Let's first consider the claim that both are truthmakers for both sentences. This falls prey to a redundancy worry—we

now have two redundant fundamental states of affairs. And redundancy is especially pernicious here, because it's explosive: on pain of arbitrariness, we will be forced to admit that highly gerrymandered but “logically equivalent” states of affairs, such as $\sim\sim\sim(\sim A \vee \sim B) \ \& \ \sim(A \ \& \ B)$, make both sentences true, and also that $A \rightarrow \sim B$ does, and so on. (And the ontological realist must distinguish between all of these states of affairs, since they all have different constituents and she is a realist about their constituents! So she can't revert to an identity claim.) So we should reject the claim that both our candidate states of affairs make both sentences true. (Some of the things I say in discussing the third response will also be relevant here.)

Perhaps *neither* of our candidate states of affairs are the fundamental truthmaker for the sentences ' $\sim A \vee \sim B$ ' and ' $\sim(A \ \& \ B)$ '. Presumably, given our background assumptions, these sentences *do* have a fundamental truthmaker. But this immediately pushes us into E-territory: wouldn't the most fundamental theory of the world involve a *single* sentence that matches the structure of whatever the most fundamental truthmaker is for ' $\sim A \vee \sim B$ ' and ' $\sim(A \ \& \ B)$ '? Our more fundamental state of affairs ought to have a more perspicuous description than either of these sentences. But that means that, if there is a single fundamental truthmaker for both ' $\sim A \vee \sim B$ ' and ' $\sim(A \ \& \ B)$ ', then *neither* T nor T' is fundamental, for the fundamental theory, T", would have a sentence the structure of which exactly matches the structure of its truthmaker.

The remaining response is also the most natural: the most fundamental truthmaker for ' $\sim A \vee \sim B$ ' is the state of affairs $\sim A \vee \sim B$, and the most fundamental truthmaker for ' $\sim(A \ \& \ B)$ ' is $\sim(A \ \& \ B)$. But we're now faced with an immediate problem: why do the states of affairs $\sim(A \ \& \ B)$ and $\sim A \vee \sim B$ necessarily co-obtain? Remember that we are considering fundamental states of affairs here. So *weak non-redundancy* applies, and we need some sort of explanation for this necessity.

One might be tempted to appeal to *logical equivalence* as such an explanation. But if logical equivalence explains anything, it explains only why the truth values of ' $\sim(A \& B)$ ' and ' $\sim A \vee \sim B$ ' necessarily travel together—not why their *truthmakers* necessarily co-obtain, and are equally fundamental. Logical equivalence can serve as an explanation for why two sentences or propositions necessarily have the same truth value—but the *reason* it can serve as such an explanation is that it suggests that two sentences have the same truthmaker! We normally conceive of logical equivalence as a relationship between representational entities, not worldly ones (like states of affairs).

In the case at hand, we're exploring the claim that these two sentences have distinct truthmakers. So we can't merely claim that the states of affairs—which are worldly, non-representational, and fundamental—are logically equivalent and leave it at that. What we would need would be some kind of *metaphysical law* that connected our two fundamental states of affairs. If one wants to call such a law “logical equivalence”, one can, but this is misleading: committing to it is committing to a meaty, governing, metaphysical law that relates otherwise free-wheeling states of affairs, not something conventional or metaphysically harmless (as logical equivalence between merely representational entities might be taken to be).

It would be easier to swallow such laws if they were to, for example, tell us that whenever we have a single fundamental state of affairs we get all of its logically equivalent states of affairs “for free”: that we are *automatically* committed to its logically equivalent states of affairs as derivative entities. But this is not nearly strong enough. It only tells us that *if* there is a fundamental state of affairs $\sim(A \& \sim B)$, then there is a nonfundamental state of affairs $\sim A \vee \sim B$. That's not what we need to explain. I've shown that the only way for **A** to be true is if *both* states of affairs are fundamental (since they are each the most fundamental truthmaker for a

sentence in a fundamental theory). So we need our law to explain necessary connections between *fundamental* states of affairs. And such a law is one that I very much doubt many people will want to commit to, and which suffers from the problems discussed in section 1.2.

Alternatively, we might want to claim that it is somehow part of, or follows from, the essences of conjunction and disjunction and negation that these states of affairs necessarily travel together. All the same worries I raised in section 1.2 apply here as well. We might additionally worry that essences serve to explain relationships between more and less fundamental entities or facts, or temporal relations between equally fundamental entities, but that positing them to explain massive fundamental ontological redundancy is exploiting this bit of theoretical machinery.

To sum up: the only way the ontological realist can maintain that **A** is true, given our assumptions about states of affairs and truthmaking, is by claiming that *both* $\sim(A \ \& \ B)$ and $\sim A \vee \sim B$ are fundamental states of affairs. But then *weak non-redundancy* applies, and we are faced with the same issues we were in the case of relations. Hence, **A** cannot be true. Since the ideological realist cannot believe **B** and cannot believe **C**, she is forced to believe either **D** or **E**, which are just *privileged* and *unfamiliar*.

I won't do so here, but it is fairly straightforward to see how we could reconstruct this story without talking of states of affairs at all. What is somewhat less straightforward is reconstructing it for the ideological realist. I do this in section 2.3.

2.3 Ideological Realism and Weak Non-Redundancy

Suppose that instead of commitment to entities to which the constants referred, that we were talking about candidate fundamental *truths*. I have stated **A-E** in terms that either the ontological

or ideological realist can accept, so there is no need to retranslate here as I did in the case of relations.

Supposing that we accept $\text{Fun}(T)$ and that ' $\sim A \vee \sim B$ ' is a sentence in T , then, if we also accept $\text{Fun}(T')$, we are committed to both ' $\sim A \vee \sim B$ ' being a fundamental truth, and ' $\sim(A \& B)$ ' being a fundamental truth. This is because T' , by definition, contains those sentences expressible in its language that are logically equivalent to those T contains.

The very same issue comes up with **A** as does when we move to nominalism in the case of relations: For the *ideological* realist, there are multiple disambiguations of **A**. First, it could be the case that the world is structured in two distinct ways, each of which is most perspicuously reflected by these two truths respectively. Second, it could be the case that the world is structured in a single way, and, as in the case of relations, these two truths are equally best ways to represent that structure, and there is no more perspicuous way to represent it.

In the former case, it is easy to see how *weak non-redundancy* applies: even if we don't want to commit to those bits of worldly structure being ontological, as soon as we ideologically commit to the idea that there are multiple fundamental ways the world is structured, we can ask why those bits of structure always “travel together”, or, in ideological realist-friendly terms, why it is that ' $\sim A \vee \sim B$ ' is true whenever ' $\sim(A \& B)$ ' is, and vice-versa, given that they can't both perfectly perspicuously describe the same structure.

In the latter case, as in the case of relations, **A** just collapses into the same position as **E**: there is a single underlying “way the world is”, or bit of worldly structure. It is just that we are additionally committed, in this case, to the idea that ' $\sim A \vee \sim B$ ' and ' $\sim(A \& B)$ ' are *tied* as being the best ways to represent this single way the world is. (Note that in this case, unlike the case of relations, there will inevitably be a large number of other sentences that will also tie with these

two, e.g. ' $\sim A \rightarrow B$ '.) But here the case is even stronger than in the case of relations that this means that the underlying way the world is *neither* resembles the sentential structure of ' $\sim A \vee \sim B$ ' *nor* the sentential structure of ' $\sim(A \& B)$ '. The case is stronger because the logical *forms* of these sentences are radically different. Neither of these sentences are perspicuous at all. Since they have two radically distinct structures but are representing the exact same bit of worldly structure, it cannot be that *either* of them mirrors that structure in a perspicuous way. For if ' $\sim A \vee \sim B$ ' was perfectly perspicuous, it would follow that ' $\sim(A \& B)$ ' was radically unperspicuous; but then it would be the case that the two sentences were not in fact tied as best representations of the world, as one would be perspicuous and the other would not. So it must be that the structure of neither of these two sentences resembles the underlying way the world is.

I've shown that one way for the ideological logical realist to endorse **A** turns out to fall prey to *weak non-redundancy*, and the second turns into something compatible with the conclusion I am arguing for: that there is just a single underlying structure with two equally good ways to refer to it (and no better way). This is much more closely aligned with the spirit of **E** (*unfamiliar*) than it is with **A**. Since it is clear how *non-arbitrariness* applies to the ideological realist, I conclude that both the ideological and ontological realist's only options are to embrace either *privileged* or *unfamiliar*.

3. What Should we Conclude?

I have argued that both the ontological and the ideological logical realist are forced to embrace either *privileged*, the view that a privileged subcollection of our logical constants is fundamental, but we can't know which ones, or *unfamiliar*, the view that the logical structure of the world is deeply unfamiliar: that it looks nothing like the logical constants that we are familiar with. This

suggests that the kind of project that some metaphysicians are now engaging in—of trying to articulate a perfectly fundamental theory of the world—is a problematic one.

There are two central upshots here. First, the obvious: either there is something about fundamental reality that is completely unknowable (if *privileged* is true), or we're working with an entirely wrong conception of what logic is like. Second, if we can't formulate our most fundamental theories without the use of logical constants, then we can't formulate an absolutely fundamental theory at all, at least not without doing a lot of work to elaborate a positive view to flesh out *unfamiliar*. I am skeptical that we can do so in a way that allows us to actually state a perfectly fundamental theory—so it may well be that even *unfamiliar* entails unknowability about the fundamental structure of the world. But we also ought not despair if we *can't* articulate such an account. There are other ways to make philosophical progress, for example, by *identifying* the irreducibly non-fundamental bits of our theories, replacing them with fundamentalia when possible, and clarifying their function (and why we think they are non-fundamental!) when not possible.

It is not just metaphysicians who need to be careful here. Philosophers of science, albeit sometimes in a different way, are also engaged in a project of trying to articulate a single best theory of the world. What is this different way? Crucial to the philosophy of science is the question of how much to *read off* our theories: what do our theories metaphysically commit us to? And intimately related: when should think that two scientific theories are *equivalent*--when they say the same thing about the world? Whether logical realism is true or not will clearly inform this question: if the purported disagreement between two theories has its roots in logical notions, and logical *anti-realism* is true, then those two theories are equivalent. But if logical realism is true, then whether the two theories are equivalent turns on the specifics of exactly

what kind of logical realism is true.

Though it's not my goal here to argue that *unfamiliar* is the better of our two views, the question of equivalence provides us with a reason for believing it is. *Privileged* has the consequence that the world discriminates between what is intuitively the same theory formulated first using '&', '~', and 'E', and second using 'v', '~', and 'A'. For *privileged* entails that some small subcollection of our logical constants is fundamental. If '&' is in that subcollection and 'v' is not, then there is an important metaphysical distinction between these two theories—we just can't know which is better. Most of us will want to reject that there is a genuine difference between these two theories. But we might not want to reject logical realism altogether. *Unfamiliar* allows us to endorse logical realism without the undesirable result that theories that seem to be mere notational variants are in fact metaphysically distinct. For it posits some unfamiliar structure that serves to anchor or ground our familiar logical constants. The proponent of *unfamiliar* can hold that these two theories that seem to us to be mere notational variants are genuinely mere notational variants, since she can hold that they both make use of logical constants that are grounded in a common, more fundamental, logical structure. What she *can't* claim is that either of these theories is *perfectly fundamental*.

Views like *privileged* are generally problematic. But should we think, in situations where we arrive at only *privileged*-like options, that the most likely explanation of us being in such a position is actually that there is some more fundamental explanation for its being the case that we are faced with options that it would be arbitrary to choose between? Should we *automatically* move to endorsing the *unfamiliar*-like option in such a case?

I suspect the answer is that avoiding commitment to unknowability claims is a theoretical virtue to be traded off against others. In particular, one might think that it is only worth moving

to *unfamiliar*-like options if we have some idea of how to flesh them out in a coherent fashion. So, return again to Benacerraf: the claim that numbers are not sets might be justified by the demonstration that it would be arbitrary for the number seven to be one candidate set vs. another; but only, perhaps, in conjunction with us being able to articulate *some* alternative view about what numbers are, if not sets. And if this is right, it's not so clear, in the case of logic, that we are justified in moving from *privileged* to *unfamiliar* without being able to say at least a bit more about our more fundamental entity or structure is.

To flesh this out, think about a somewhat similar example. Perhaps, Newton was justified in holding his version of substantivalism about space despite it being the case that Leibniz argues that this violates the principle of the identity of indiscernibles. We might characterize the issue as one about knowledge rather than the PII: perhaps what Leibniz showed is that there's something fundamentally unknowable about the world if substantivalism is true. Why might Newton have still been justified in believing in absolute space in the face of this challenge? He was forced into the same sort of situation we are here—of choosing between a position very much like *privileged* (substantivalism) and one very much like *unfamiliar* (since he rejected relationalism as a candidate replacement for substantivalism).

Newton's bucket argument seemed to show that the only available way of fleshing out his unfamiliar-like option—relationalism about space—failed. If there was no way for the relationalist to account for the bucket argument, then one might think: unknowability and a proliferation of possibilities is superior to not being able to account for the empirical phenomena. If there was no satisfactory relationalist theory, and there was no alternative to Newtonian substantivalism and relationalism, then Newton's position was a reasonable one.³⁴

³⁴ I am not suggesting that this picture is necessarily exactly historically accurate, and it shouldn't matter too much whether my reader objects to this broad outline of an incredibly difficult debate. If you do object, just treat the case as presented as a toy one. The point stands. For discussion of this point, see Dasgupta (forthcoming),

Newton was, perhaps, justified in endorsing a position that looked like *privileged*, because he believed that there simply was no acceptable candidate more fundamental theory one could move to.³⁵ This is a tricky issue. One might still hold that he was *not* so justified, because being in a *privileged*-like situation at the least presents us with a demand to take seriously a search for some more fundamental grounds for our theories.

The point here is that the important question in our case is whether there is more we can say about *unfamiliar*—that is, whether we can familiarize ourselves with some of the unfamiliar options. I think that it is both unclear whether this is possible, and incumbent upon the logical realist to try. This is a point of methodological disagreement between Sider (2011) and myself. His view is that we should accept the proliferation of *actual* logical constants due to a lack of a more fundamental theory to move to. I have argued against such a view here, but I have not extensively argued against the unknowability claim, *privileged*, nor have I put forward any sort of positive view about how to make *unfamiliar* more familiar. So, while I find *unfamiliar* more appealing than *privileged*, more work has to be done to justify moving from the latter to the former.

In section 1, I argued that we are forced to believe either that we can't know which of a pair of converse asymmetric relations is fundamental, or that neither is. What might replace a pair of relations as a single fundamental entity or bit of worldly structure that would eliminate this problem? We might look for single symmetrical relations that could somehow do the job (perhaps with some additional theoretical apparatus). This may well be harder than it sounds, but

Maudlin (1993, 2012), Sklar (1974).

35 Whether he was right to think that there was no such theory is unclear, especially given advances in the philosophical literature on this dispute. Sklar (1974) points out that the relationalist *can* give an acceptable explanation of the bucket phenomenon. But that theory is both *indeterministic* and *non-local*. And at this point, we are simply trading theoretical costs against one another.

at least there are suggestions for how to do it.³⁶ With logic, on the other hand, we are left at something of a loss when looking for a more fundamental theory to ground the logical constants.

We might try to appeal to a primitive entailment relation as the more fundamental entity that the constants are derivative of. But there are problems with this strategy. First, it leaves us with a sort of mysticism about what this entailment relation is *and* about how it generates the logical constants that we know and love. We can see this just by thinking about how accounts of entailment either always seem to reference the constants qua truth-functions, and hence would be badly circular if we were trying to claim that entailment was fundamental, or appeal to purely proof-theoretic notions, which are far more friendly to logical anti-realism than realism to begin with. Put more simply, entailment is at least conceptually explicated in terms of other notions that seem to be *conceptually* prior. This does not, of course, mean that entailment is *metaphysically* prior to these notions, but it does put the burden on us to give an account of how entailment can ground e.g. the constants, and why the conceptual and metaphysical orders of explanation are reversed.

Second, note that entailment itself is an asymmetric relation of the sort that the first part of this paper raised problems for. The problem for logical constants and the problem for asymmetric relations are not just analogous—they are in this sense intimately entwined. If we are going to move to taking entailment as the primitive, fundamental entity or structure, much more work needs to be done to flesh out what that would look like, and how this issue would be avoided.

There are of course other strategies we might take. Elsewhere, I develop an analogy between quantitative predicates and logical constants. Quantitative predicates face us with

36 e.g. in Dorr (2004), Fine (2000), Williamson (1985) (perhaps).

similar issues to the ones outlined here. Regardless of what one's view is about the underlying metaphysics of mass, for example, one should think that a theory that makes use of the mass unit the *gram* contains something non-fundamental. For we should either think that there is a privileged mass unit, but that we don't know what it is, or we should think that there is no privileged mass unit. And this means that some of the choices we make about our mass statements in our theories are arbitrary: if a *schram* = 1.4 grams, then it is arbitrary that our theory claims that some object has mass 1.4 grams, for there's no reason to think this is a more fundamental description of reality than that that object has mass 1 schram. The analogy here is fruitful because the issues are clearer, and the methods for dealing with them better developed, in the case of mass. Measurement theory provides us with for systematizing and illuminating what are in some sense very similar issues to those logic faces us with. We can develop some of these tools to deal with the logic case. But this doesn't give us a *positive* account of what the fundamental logical structure of the world is—of how to familiarize ourselves with the unfamiliar. The hope, though, is that it will clarify what the *constraints* are on the unfamiliar, by identifying commonalities in structure between theories that are equally good at representing reality.

And surely there are other ways we might try to flesh out *unfamiliar*, as well as serious challenges that both *privileged* and *unfamiliar* face us with. But I won't elaborate further on them here. I have argued that logical realism commits us to one of two surprising views: that the logical structure of the world is unknowable, or that it is deeply unfamiliar. These views have serious consequences for our philosophical and scientific theorizing, and we ought to consider them carefully.

3. Radical Representation and Logic as Triadic

When we teach undergraduates logic, and they are introduced to soundness and completeness proofs, they often wonder why soundness and completeness *matter*. Haven't we just designed our system so that our proof theory and our model theory align? How is it then philosophically meaningful or important to show that our system is sound and complete? While there is some sense in which these questions are born out of a misunderstanding, there is a deeper sense in which they are very good questions, and they are alive and well today in the philosophy of logic.³⁷

In this paper I present a novel answer to the question of why soundness and completeness matter. I argue that the *logical realist*, who thinks there is logical structure *in the world*, should think that the syntax and semantics of logic together aim to *represent* that logical structure. She should think that logic is triadic: that it consists of a syntax, a semantics, and a *target*, and that the world itself, or the logical structure of the world, is the target of the “one true” logic. I flesh out this claim by giving an account of *radical representation theorems*, which, as more familiar representation theorems do, establish a kind of structural isomorphism between multiple representations of a phenomenon. I then show that soundness and completeness together form a radical representation theorem: we can perform *transformations* between syntax and semantics without losing information. I argue that this is what justifies, on the triadic view, the claim that syntax and semantics represent the very same target.

In section one, I'll lay out some assumptions about, and basic motivations for, logical realism. I'll also motivate a picture of logic as *triadic*: as a relationship between syntax,

³⁷ See, e.g., Beall and Restall (2013), also Dresner (2010) for a discussion of the relationship between the question of the philosophical importance of soundness and completeness and questions about logical consequence.

semantics, and a *target*. The natural combination of these two views is to think that the target of the “one true” logic just *is* the logical structure of the world itself. In section two, I'll give a general account of *radical representation theorems*, and sketch the reasons for thinking that soundness and completeness together form one. In section three, I'll lay out and discuss some potential upshots of the view

1. Logical Realism and Logic as Triadic

Generally speaking, a logical system can be viewed from (at least) two equally important but fundamentally different angles: i) it can be viewed as a pair formed by a syntax, i.e. a deductive system, and a semantics, i.e. a class of mathematical structures onto which the underlying language is interpreted, or ii) it can be viewed as a triad consisting of a syntax, a semantics and the target phenomenon that the logic is intended to capture. In the first case, both syntax and semantics are viewed as autonomous mathematical structures, not owing anything to external elements. In the second case, both syntax and semantics are accountable towards the primitive target phenomenon, which may be an informally formulated concept, or even phenomena in the 'real world' (e.g. logics of action, logics of social interaction, quantum logic etc.) Indeed, in the second case, both syntax and semantics seek to be a 'model' in some sense or another of the target phenomenon. (Dutilh Novaes and Andrade-Lotero 2012 p. 388.)

Dutilh Novaes and Andrade-Lotero claim that the above is “perfectly uncontroversial”. I am sure nothing is perfectly controversial in philosophy, but this seems a good place to start. If we view logic in the first sense outlined above (that is, as a *pair* of syntax and semantics), all soundness and completeness tell us is that there is a relationship between our two autonomous mathematical structures. But, I'll argue, the logical realist ought to think of logic in the second, *triadic* sense. And this will also help answer our students' question about the importance of soundness and completeness: if we are viewing a logic as triadic, then the significance of

soundness and completeness is that they show us that, if one of syntax and semantics represents a given target, then so does the other.

I'm going to *assume*, rather than argue for, the claim that logic should be thought of, at least sometimes, as triadic. Of course, we use logic for different purposes, and there may well be cases in which we are best served by viewing logic as dyadic. But it's worth thinking about what kinds of things might be targets on the triadic conception of logic. As Dutilh Novaes and Andrade-Lotero point out, some potential targets will be more *conceptual*: we will be trying to represent a concept we have. Others will be more *worldly*: we will be trying to represent something that we take to be, in some sense, mind-independent and part of the nature of the world. My focus will be on the second kind of target. But the first is more familiar, so I will start by briefly saying something about it.

Kreisel's (1972) well-known “squeezing” argument, which attempts to triangulate between the syntax and the semantics of first order logic, and our *intuitive concept of logical validity*, assumes that logic is triadic, and that it consists of two distinct representations of a single target, that is, our concept of logical validity. The idea behind the squeezing argument is that we can show that there is an important relationship between provability, truth in every model, and our intuitive notion of logical validity: that they all align. In order to do this, we need to appeal to completeness. This despite the fact that our intuitive notion of logical validity is an *informal* notion whereas provability and truth in every model are *formal* notions. So we can, in some sense (assuming Kreisel's argument is successful) use completeness to help show that both “halves” of our formal systems are getting at the same kind of informal system or intuitions.

The Kreisel argument is a bit too long-winded to rehash here, but let me just briefly outline Field's (1980) adaption of it. Suppose that we have a conception of logical truth (or

logical consequence, but let's stick with truth). In one sense the idea is very simple: if that concept of logical truth is *intuitively sound* and *intuitively complete*—that is, if P is provable, then it is logically true, and if there is a countermodel to P, then P isn't logically true (that is, there's no “counterinterpretation” in our conception of logical truth). Then, it follows that if our system is complete, all these notions—truth on every interpretation, provability, and logical truth—are coextensive.

The details aren't terribly important to this paper. My point here is just to show that logic is commonly conceived of in this triadic way, and that there is already at least one argument that is supposed to show something about the relationship between logic's *target* (in the Kreisel/Field case, our concept of *logical truth* or *logical consequence*) and the two representations (syntactic and semantic) of that target.

But as Turner (2016, introduction) clearly points out, it is important not to read anything extra into Kreisel's argument. It's not an argument that is meant to tell us anything at all about the substance of our target notion, in Kreisel's case, our intuitive notion of logical validity. Rather, we *presuppose* the target notion and then show that there are some syntactic and semantic systems that (a) align with one another and (b) adequately reflect the target notion. And even this makes the argument sound more magical than it really is: the reason the argument works is that we assume a close relationship between the target notion and both our proof theory and model theory: that our inference rules are sound with respect to the informal target notion, and that if we've got a model-theoretic counterexample to an argument, we've got an intuitive counterexample as well. Squeezing isn't supposed to *reveal* anything about the nature of our intuitive concept of validity to us, rather, we need that concept of validity to be fleshed out enough that we can say enough things about it that we can, in fact, generate the necessary

premises to “squeeze” it between proof-theoretic and model-theoretic validity.

Kreisel's argument—and various modifications of it, e.g. in Etchemendy (1990)--concern our intuitive concept of logical validity. Here, I am not concerned with our intuitions about logical validity. Because the thing I want to consider, as the target of our triad, falls under the category of “phenomena in the world”. How might worldly phenomena be the target of representation of a logic? A logic of action might take *action* as its target. Many of us are familiar with the development of and philosophical discussion surrounding *quantum logic*. It's hard to see what it would be to take quantum logic philosophically seriously without having a view of logic as triadic and of the *world*, or at least, the quantum world, as the target of quantum logic. Deontic logic clearly has a target, though we might disagree about exactly what that target *is* and even perhaps whether it is worldly or conceptual. (Perhaps at least some moral realists ought to take it to be a worldly target, whereas moral anti-realists might be more inclined to take its target to be purely conceptual. Of course, there is space for some intermediate views here as well.)

I assume that *logical realism* is true: that there is something about the ontology, or structure, of the world that our logic, if successful, reflects or captures. The attempt to develop and adopt quantum logic was a project of *determining what logic best fit the structure of the physical world*, and hence, involved a worldly target (the world, or the quantum world). The logical realist thinks, at the least, that there are better and worse logics for reasoning about and theorizing about the world, and, (the controversial bit), there is something about the *mind-independent* world that *makes* those logics better and worse for theorizing about the world. So it's not just that the structure of our thought requires us to reason using a certain logic, or that pragmatic considerations constrain our choices of logic. Rather, the idea is that the world is

structured a certain way, and that a logic can be better or worse at capturing that structure.

What does this mean? It's very hard to say. The easiest way into logical realism is, I think, to consider an analogy with the famous case of the predicates 'green', 'blue', 'grue', and 'bleen'. We can use either 'grue' and 'bleen' or 'green' and 'blue' to describe the world truly; but many of us think that there is something better—and not just pragmatically, but also *metaphysically* better—about using 'green' and 'blue' to do so. 'Green' and 'blue' better respect the *structure* of the world, and are hence more *metaphysically perspicuous* than 'grue' and 'bleen' are. But note that we can ask this same question about logical constants: Are '&' and '~' perfectly perspicuous? 'v' and '~'? '↓' (the 'neither/nor' connective)? What about quantifiers?

If the world has any kind of *structure* (perhaps *color* structure) that 'grue', 'bleen', 'green', and 'blue' correspond better or worse too, then the question of which of these predicates are more or less metaphysically perspicuous makes sense and has an answer. If the world has logical structure—if *logical realism* is true—then these questions must make sense, and must have answers. Now none of this is to say *what that structure is like*—it merely tells us one commitment we might have in light of being realists about that structure. But I think it is helpful for gaining some understanding of what it is to commit to there being logical structure in the world. (And note: just as we needn't commit to there being worldly *properties* of greenness and blueness to think that 'green' and 'blue' are more metaphysically perspicuous than 'grue' and 'bleen', we also needn't commit to the idea that there are actual *entities* that our logical constants represent to be logical realists in this sense—see 'Following Logical Realism Where it Leads' for more detail here.)

Elsewhere, I give a conditional argument for logical realism by showing that *metaphysical* realists must also be logical realists. I show that, for example, if logical realism is

false, then we can't distinguish between a view on which the world is fundamentally made up of both individuals and properties that those individuals bear, and a view on which the world is fundamentally made up only of properties.³⁸ While his argument is quite different, Sider (2009), (2011) argues for a form of realism (that is, for the claim that the world has *quantificational structure*), in response to similar deflationary challenges to distinguishing between various views of *composition* and *persistence*. I won't here rehearse either of those arguments, but will merely suggest that they at least should give the metaphysical realist pause about immediately rejecting logical realism.

One might also be moved by the following consideration: I would venture that most philosophers are not logical pluralists, or if they are, are narrow logical pluralists. We tend to think that there is *one true* logic (or perhaps multiple true logics, but not that anything goes), and, even if one doesn't think that there is logical structure that looks just like that logic in the world, one needs to identify *what it is that makes this the case*. One answer (which I'm not endorsing) is something like “norms of rationality”. But what grounds these norms? Why ought we reason according to a particular logic?

On the picture of logic under consideration, if we think there are *no* such norms—no constraints on how we should logically reason--then we must be very permissive logical pluralists. If, instead, we want to claim that there are constraints on how we should logically reason, we're left with two options: these constraints are somehow internal to, or come from, us, or they are external to us, or at least are grounded in something external to us. If one has that kind of view, one is still, I think, a logical realist: one thinks that there is something about the mind independent world that constrains the way we should logically reason, and hence that

³⁸ See the introduction of this dissertation for a slightly more expanded summary of how this kind of argument goes.

makes it the case that a particular logic is correct. Whether this version of logical realism is strong enough to move one to endorse the idea argued for in this paper will depend on the particular details of the view, but hopefully it provides some motivation for taking the picture seriously: if we are going to be believers in “one true” logic, we need a *reason* for believing, and believing in mind-independent logical structure gives us a very straightforward justification.

Again, I have not argued for the triadic conception of logic, nor, really, for logical realism. But I have sketched some motivation for each of them. And with these two assumptions on the table—that the triadic conception of logic is (at least sometimes) correct, and that some weak form of logical realism is true—we can conclude that, at least some of the time, the *target* of our logic is just the world, or the structure of the world, itself. In addition, logical realism itself helps *motivate* the idea that the two sides of logic—proof theory and model theory—are *representational*, and hence helps motivate the triadic conception of logic. For if logical realism is true, then there is something external to the syntax and semantics of our logic and language that that syntax and semantics is *aiming at* capturing: the logical structure of the world itself. Philosophers of logic will probably be more comfortable with this idea of proof theory and model theory as representational than metaphysicians are (though perhaps not with the logical realism!), but I hope to have given some motivation for thinking of proof theory and model theory as their own representational systems.

In the following section, I will flesh out this notion of syntax and semantics *representing* their target. I will do so by giving an account of *Radical Representation Theorems* ('RRTs'), and arguing that soundness and completeness together form an RRT. RRTs are like their more familiar cousins (traditional representation theorems) in that they show a kind of structural isomorphism between representations. This is what *justifies* a belief that both sides of a logic

successfully represent the very same target. It also gives us a more general way to think about representation when we lack access to the target of representation (e.g. when the target is an abstract object, or when we cannot empirically observe it).

2. Radical Representation Theorems and Soundness and Completeness

What is it that makes it the case that, for a given logic, that logic's syntax and semantics successfully represent *the same* target? First, there needs to be some target that one of them successfully represent; and second, syntax and semantics need to be structurally isomorphic (in some loose sense, at least), in order to show that, if one side successfully represents a particular target, so does the other. So there are two questions we can ask generally on the triadic picture of logic: first, whether *either* side of the logic successfully represents a particular target (or *any* target at all), and second, whether, if that is the case, so does the other. And there are two more specific questions the logical realist can ask: first, whether either side of the logic successfully represents the logical structure of the world; and second, whether, if that is the case, so does the other.

Now, as I argued in the first paper, I think it's quite hard for us to get an answer to the first question. It's at least, I think, impossible to *directly* answer this question. But we can, I think, answer the second question. And it is worth doing so for a few reasons. One is that doing so might help give some insight into the first question *via* an answer to the second question. A second is simply that it will give us a clearer picture of the triadic conception of logic, and particularly what it should look like from the perspective of the logical realist.³⁹

³⁹ A third is that it might help us begin to develop some answers, for the logical realist, to certain questions in the philosophy of logic, for example, to what the nature of logical constants is. This would have to be developed much further, but the idea is that accounts of the logical constants tend to take either proof theory or semantics as primary in telling us what logical constants are; on the triadic, realist conception of logic, this is the wrong sort of route to go down, since an account of the constants should somehow bottom out in our *target* and not

To answer the question of when syntax and semantics successfully represent the same target (assuming one or the other does), I'm going to give an account of *Radical Representation Theorems* ('RRTs'). Before I do that, it will be useful to think briefly about representation theorems more generally.

Representation theorems are used in measurement theory along with uniqueness theorems to show that a given kind of representation of a quantitative structure preserves the structural features of that quantitative structure. Very roughly, representation theorems just establish that there is some mapping from the target domain (the quantitative structure) to the domain posited by the representation (numbers), and uniqueness theorems tell us what kinds of transformations preserve that mapping.

In a bit more detail: First, in order to establish the right kind of relationship between the structure of our target domain (D) and the structure of our surrogate (represented) domain (D'), we need to show that there exist *homomorphisms* from the relations on the target domain to the relations on the surrogate domain. A homomorphism is just a structure-preserving function that maps every member of D to a member of D'. If our D' contains the real numbers, and our D contains physical objects, the idea is that we can select, e.g., a homomorphism that preserves all the mass ratios that hold between objects in D by mapping those ratios to the equivalent ratios on D'. So if the coffee in my glass gets mapped to the real number 3.7, and the water in my water glass is in fact twice as massive as the coffee in my coffee cup, the water in my water glass will get mapped to the real number 7.4. *Representation theorems* show that there exist such homomorphisms from D to D'.

A *radical* representation theorem is like a representation theorem in that it is structure

either representational side of logic.

preserving in a certain way: the hope is that it, like a normal representation theorem, shows us that there is a certain kind of structural isomorphism between two entities. But in the case of a normal representation theorem, we have a clear, empirically observable *target*, and what we are showing is that a certain kind of structure that target has is *preserved* by a given representation of it. In the case of radical representation, we lack any access to that target that is independent of the very representations we are evaluating. So, for example, the target might be an abstract object, or some other kind of unobservable entity. In such cases, we can still establish a kind of structural isomorphism *between representations*. However, it is going to be a bit messier what *counts* as such a structural isomorphism, since it won't be simply a mathematical result, and also a bit messier *how we can know* whether there is such a structural isomorphism. This is in part because it is going to be far less clear what counts as the *domain* of the representation. In this sense it is a bit of a cheat to call RRTs *representation* theorems. But I will say more, in what follows, that hopefully will persuade my reader that there is some relation between normal representation theorems and their more radical cousins.

One way that we might begin to answer the question of how we can know when we have this kind of structure-preserving relationship is by showing that there is a translation procedure for going back and forth between two representations, and that that translation procedure preserves important structural features of each representation. When we can do this, we (generally) have an RRT. The proposal will become more clear as I go on, but it's important to note that the translation procedure itself is just a rough guide to there being the right kind of structural relationship between two representations, and that it will overgeneralize—not every possibly translation procedure that preserves *some* structure should count as an RRT. I will say more about this, and why it is less problematic than it sounds, at the end of this section.

To understand radical representation, it is easiest if we first suppose that there are abstract objects like symphonies and triangles. We needn't remain committed to this idea to see how soundness and completeness might serve as an RRT; we can be logical realists without being committed to the existence of abstract objects (both Sider (2011, ch 10) and I (in the first paper of this dissertation) show that there are coherent forms of nominalist logical realism). So, one can reject all of the particular examples that I give in what follows, without thereby rejecting that the concept of an RRT is useful, and even without rejecting that RRTs can help us better understand representations of aspects of the world that are unobservable.

Consider the relationship between a particular performance of a symphony, and the written score for that symphony. Or the relationship between a picture of an equilateral triangle and an algebraic description of an equilateral triangle. It's clear that, for each of these pairs, there is something that their members have in common. One way—albeit, in all three cases, a controversial way—to conceive of what it is that the members of each pair have in common is that both of them *represent* the same underlying entity. So we could think of a particular performance of a symphony as one way of representing that symphony, and its written score as another. And we could think of both images of triangles and algebraic facts about triangles as both representing *the triangle*.

We have a *radical representation theorem* when we have a demonstration of a certain kind of equivalence between two different kinds of representations, that shows that they are structurally isomorphic (in a certain sense), and hence can represent the same *target*. This demonstration of equivalence will take the form of us being able to reproduce a representation of one kind from the other, and vice-versa. Sometimes that demonstration might involve something like a mathematical proof (which one might think that we have in the case of the triangle). But

often it will be much messier and harder to get a handle on, such as in the case of the symphony (assuming, for the time being, that both the written score and the performance count as representations of the symphony). What counts as a demonstration of this fact? One direction is easy: that we can *reproduce* something extremely similar to the performance if the only information we have is the score. The other direction is a bit harder: that we can reproduce something extremely similar to our written score purely from listening to a performance. But doing so does not seem to be outside the bounds of human capabilities.

Take, for example, a particular copy of the written score of, and a particular performance of, Aegri's 9-part choral work *Miserere*. As the fourteen-year-old Mozart demonstrated when, after listening to a single performance of *Miserere*, he transcribed it perfectly, it's clear that we (or at least, the most musically inclined among us) can reproduce something that looks almost exactly like our copy of the written score purely from the information we gain from listening to a particular performance of the work.

This is a prime example of an RRT. It tells us that these two radically different kinds of representations encode quite a bit of the same information. And, if we were antecedently convinced that, for example, our particular performance was an auditory representation of an abstract object, I claim, we ought to now think, once we have performed these “translations”, that our score does as well.

It's important that this translation procedure is *structure-preserving* in an important way in order to count as an RRT. It's clear that a copy of a written score of a piece of music and a particular performance of that piece of music are structurally isomorphic in some sense; the left-to-right dimension of the staff in the written score corresponds to the moving-forward-in-time dimension of the performance; the height of a given representation of a note on the staff

corresponds to the pitch of a note in the performance; the shape of a given representation of a note corresponds to how long it is held in the performance; and so on. This preservation of structure is what *enables* us in reproducing the representation of one type from the other, and vice-versa: there is a structural core that is preserved by each. In all cases of RRTs, there will be a kind of structure preservation of representations; but while perhaps in this case there is at least some loose sense in which we can produce homomorphic functions (that map, say, “notes” on the page to actual sounds, or vice-versa), we won't always be able to do this.

When it comes to representation theorems, we need information about their target before we can get the right kind of thing. In the case of mass, we have an actual mass structure in the world that we can point to: the reason we care about *these* common structural features in our representations of mass is that we think they are central to what it is to adequately represent mass structure in the world. But that's because we can *look at* the mass structure in the world and see that it shares these common structural features. All of the examples of RRTs I gave above, *if* they have targets (that is, if I'm right to call each pair a pair of 'representations' in the first place), have targets that are *abstract objects*: a particular symphony, an equilateral triangle. And it's hard to see how we can have information about the nature of these abstract objects *independent* of our representations of or concepts of them, so it's hard to see how we could get anything remotely like the relevant information about the target when it comes to radical representation. This is the central thing that separates normal representation from radical representation: in cases of radical representation, we lack any kind of independent observational access to the targets of our representations.

Once one has adopted this framework, it is fairly natural to think that *soundness* and *completeness* theorems give us something like the translation process described above between a

written score and a performance of the same piece of music. Consider the account I gave above of what we might take to count as an RRT in the case of the written score of a symphony and a particular performance of the symphony. The question was whether, given the score, we could reproduce something that sounded and looked a lot like the particular performance, and whether, given the particular performance, we could reproduce something that looked a lot like the score.

For soundness and completeness to form an RRT, they must show us something comparable to what our translation scheme between the performance of the symphony and the written score of the symphony show us: that, given a semantics/model theory for a logic, if that logic is sound, we can reconstruct some inference rules/proof theory for that logic that are going to look a lot like that logic's actual inference rules; given a set of inference rules/proof theory for a logic, if that logic is complete, we can construct a semantics/model theory that's going to look a lot like the semantics we've already got. (That is, imagine we only had access to, say, the semantics for the logic, and knew generally what kind of thing inference rules were, and that the logic was sound—then, to get an RRT, we would need to be able to produce a set of inference rules that looked a lot like the *actual* inference rules for that logic. And likewise for completeness.)

Let's start with considering what for these purposes is the simplest case: axiomatic classical first-order propositional logic, with the horseshoe as its only connective, and a single inference rule, modus ponens. In such a system, if we know only that the system is sound and complete, then, first, if we only have access to the facts on the 'semantic' side of the system—the relevant one being what the truth table for the horseshoe looks like—there is only one non-trivial inference rule that we can possibly construct: modus ponens. And, if we only have access to the facts on the proof-theoretic side of the system—that is, that modus ponens is our single inference

rule—there is only one truth table for the horseshoe we can possibly construct.⁴⁰

One might have worries about the semantics to inference rules side of the picture. Supposing, for example, that we have a symbol, 'T', for tautologies, we could construct the following rule ('tautologized modus ponens' or 'TMP'): $(T \supset P) \supset Q$, $T \supset P$, therefore P . This rule is equivalent, in a strong sense, to modus ponens. But we can construct *either* this rule or regular old modus ponens from the semantics for a language that has the horseshoe and the 'T' symbol. Now, if we simply claimed that TMP and modus ponens *were the very same inference rule*, then there would be no worry here: we are simply, perhaps, faced with producing two different variations of the same inference rule, and it doesn't matter which we produce (via our translation procedure). Should we worry about the fact that even introducing something as simple as the T symbol to our language makes our translation procedure less strong (since we have more options for which inference rules we can generate from our semantics)?

It is unclear whether modus ponens and TMP are the same inference rule. But that's because we're working with a picture of logic on which we don't know everything about the target that our proof theory and model theory are aiming to represent, and the individuation conditions for inference rules are going to be at least partly determined by the nature of that target. That is, the reason that it's hard for us to say whether modus ponens and TMP are the same rule is that we don't know *what it is that would make them the same rule*. When we're trying to evaluate a candidate RRT, and don't know the nature of its target, we won't have an objective measure of whether two representations, even two representations of a single kind (say,

40 The latter of these two claims is easily provable by cases (showing that each alternative possible truth table for the horseshoe gets us a mismatch between provability and truth in all models). The other direction isn't as simple, since there are infinitely many potential inference rules, but I'll just briefly sketch the way a proof would go: first, it's easy (and familiar) to show that modus ponens “matches” the truth table for the horseshoe in the right way. Next, we would need to show that all of the cases of potential inference rules are instances of a finite number of general forms that the rules could take, and show that none of these general forms (except modus ponens) gets things right.

rules of inference) are equivalent. So what should we do?

One solution is to introduce something analogous to a *uniqueness theorem*. In measurement theory, uniqueness theorems show that our adequate representational systems all bear strong relationships to one another. Here, we will have something like this purely in our proof theory: in cases where we have multiple possible inference rules that “match” our semantics, we can demonstrate a very close relationship, of interchangeability and interderivability, between such candidate rules. So, in this case, we would have something like the following: while we might generate an *identical* inference rule from a semantics (or vice-versa), any inference rule we could generate would be interchangeable and interderivable with the inference rule that our system started out with.⁴¹

In this simple case, soundness and completeness form a nearly, but likely not entirely (because there might be multiple candidate inference rules on the proof-theoretic side), *perfect* RRT, where a perfect RRT is one where we have a way of translating between the two representations that recovers *all of* the features of the representations in both directions. RRTs don't have to be perfect when they show us how to move between one representation and another. What matters is that they share a generous core, and that we have something like uniqueness theorems that relate, on one side of the RRT, the different possibilities that we might *recover* from the other side alone. It is clear that modus ponens and TMP are, at the least, *incredibly* similar inference rules, and that it is easy to show that, even by the lights of proof theory alone, they are extremely structurally similar. That's good enough, for our purposes, to get a useful RRT.⁴²

41 I think there are some subtle issues here, but I don't have space to expound upon them here.

42 Of course, we can also show that modus ponens and TMP have the same truth table. But it would be circular to use that as a criterion of identity, or even similarity, for derivation rules when evaluating soundness and completeness for being an RRT. We can't assume that *having the same truth table* is sufficient for *being the same inference rule* if our goal is to determine whether we can “read off” inference rules from semantics and

In this simple case, soundness and completeness form a near perfect RRT. As soon as we move to more complex systems, though, things will get messy much more quickly. Still, I think we can apply what I've just said to more complex systems.

Before I move on to, in section three, talking about potential upshots of thinking of soundness and completeness as together forming an RRT, I want to discuss some potential *epistemic* objections to what I have said so far.

First, how far can RRTs take us? In the cases I described above, while it might be controversial whether or not the members of our pairs were really *representations*, or were representations of the *same target*, it is not all that controversial that there is important structure that the members of each pair share, and quite a bit of it. This is why it seems that we are justified in thinking that if one member of the pair represents a particular target, so does the other. One might think, though, that the concept of a RRT is to do much work for us, it better not be horribly vague what counts as an RRT, and that it better not be the case that any old equivalence will do. Now, in the case of a normal representation theorem, say, about mass, we don't have to worry—at least, not as much—about what “counts” as a representation theorem, and how we can know when we do or don't have one, because representation theorems are just mathematical results. But it's not as obviously transparent, when it comes to RRTs, what *counts* as a structure-preserving mapping and what doesn't. One can think of the translation tests I've given (can we *reproduce* a performance of a musical work merely from its score, without access to what it sounds like? Can we reproduce a drawing of an equilateral triangle merely from looking at an algebraic formula? etc.) as a kind of epistemic tool: if we can do these things, then there's a good chance that we've got the right kind of structure-preservation between the two

vice versa doing so *assumes* that soundness and completeness form an RRT, or at least, comes close to doing this.

representations in question.

I see no problem generally with having to use a loose, epistemic guide to when we have a structure-preserving mapping. Of course, mathematical results are prettier, and nicer. But it's to be expected, I think, that in the case of something like an RRT, we can't always get such a nice, clean result, and yet there is still something interesting to be said. The problem, though, is that this guide to structure-preserving mapping overgenerates. For example, the beer aficionados among us might be able to blindly taste a particular beer, and reproduce something that looks just like the label that is typically found on that beer's bottle, and perhaps upon looking at the label for a beer, might be able to brew something that tastes quite similar. This is at best, though, a borderline case of structure-preservation. Where do we draw the line as to *when* a translation procedure gives us an RRT?

There are further misguided translations we could produce. What about the taste of a beer and the label of a different beer? Even the latter pair's members have something in common: if we grant that tastes can be representations (more on this in a moment), then they seem to both represent the kind *beer*, even if they don't both represent some more fine-grained abstract object, e.g. *I.P.A.* And there is a way to create a translation-reproduction scheme that allows us to reproduce one from the other. Should this count as an RRT? Should we be worried that, pushed to the extreme, the concept of an RRT is useless if it includes such things (and worse)?

I don't think we should be worried. While it is true that I haven't given necessary and sufficient conditions for a translation-reproduction being structure-preserving in the right sort of way, it's fairly intuitive that these examples should not count. Likewise, it's fairly intuitive, in general, that mere names, like 'Michaela', while they may be representations of their referents, don't *structurally* represent those referents. Now, something's being fairly intuitive should only

carry us so far; and it seems like wherever we draw the line between something being a good *structural* representation and not will be a bit vague, and that wherever we draw the line between two representations being *structurally isomorphic* or not will be a bit vague when we can't necessarily appeal to a strictly mathematical demonstration of structural isomorphism.⁴³ Is there more to be said here?

Yes, though it will also be a bit imprecise. It is true that there is *some possible* translation procedure between the taste of one beer and the label of a different one, and so these translation procedures come too cheap to guarantee RRTs. So we want to require our translation procedure to be structure-preserving in the right kind of way. There are two requirements we should make of it to ensure this. The first is that it not be a product of *stipulation* or *memorization* of particular cases. For example, in the beer case, we can't just use a translation process in which we merely memorize that when we see the label of beer x, we should match it with the taste of beer y.

Of course, being able to translate between two very different kinds of representations *does* require some kind of memorization tasks (e.g. memorizing which placement of a note on the staff corresponds to which tone produced on an instrument), and some stipulation tasks (without us knowing for at least a couple nodes on each structure that they correspond to one another, we can't reproduce them from one another—compare representation when it comes to mass—there is a sense in which we *stipulate* that an object in the world has mass one gram, and without doing so can't apply the gram scale to mass-in-the-world). But these memorization and stipulation tasks are *general* and not *specific*. What I mean by that is that we might stipulate some minimal things about the relationship between nodes in a structural representation of kind

43 Though I note that the more formal notions here are also often *philosophically* slippery in the same sort of way.

K and nodes in a structural representation of kind K'; but doing so will allow us to translate between *any* representation of kind K and the corresponding representation of kind K', and vice-versa. So whatever stipulation and memorization tasks we need to accomplish, they have to allow us to *generalize* to all cases of representation of the kinds involved.

One way we might have a translation procedure in the case of a label of one beer and the taste of another would be to memorize that label x corresponds to taste y; but this memorization/stipulation isn't generalizable—it doesn't help us translate between any other labels and any other tastes. So this way of producing a translation is blocked by the requirement that we should only allow (as guides to RRTS) translation procedures that employ generalizable memorization and stipulation.

But another way that we might generate a translation procedure between the label of one beer and the taste of another would be to use our metalanguage (or whatever we are using to “translate”—in the case of the musical work, it might be a bit misleading to claim that the cognitive process of going back and forth between written and aural music is conducted in a language, and it may be misleading in, e.g., the triangle case for similar reasons) to *insert* a lot of structure into the translation process. So we might have a very complicated translation procedure that maps beer labels to tastes, and this kind of complicated procedure can't necessarily be ruled out by the anti-specific-stipulation claim I made above—it might be systematic and general in the right way, without genuinely preserving the kind of structure we want to preserve. Imagine, for example, a translation procedure that tells you that if you have a representation with the property of redness, you should ignore all of its other properties, add nineteen distinct properties, and translate the resulting 20 properties via some elaborate code into musical notation. One can imagine that we could create systematic translation procedures of this

kind, and we can't necessarily rule this out with the requirement that we only have generalizable stipulation. (Because this case *is* generalizable in the right way: imagine, for example, a complicated program that had built into it a perfectly general formula for moving from any color input to a complicated list of 20 properties to be encoded into a different kind of representation.)

So we need a second kind of condition on what kinds of translations between representations count as being a guide to RRTs: it can't be that the translation involves the *insertion* of a lot of new structure into the picture (and then translation of that new structure into the second representation), or the *deletion* of a lot of the structure from the first representation when translating to the second.

In general, the more structure that our representational entities have that our target lacks, the less simple the translation process between them will be, and the less inclined I am to think that a translation between them should count as being a guide to an RRT. In many cases, including the case of logic, we don't have a completely independent picture of the target we are trying to represent. And the less 'noise'--the less extra structure we have in our representations--the more likely we are to be able to be able to *see* what our representations have in common, and hence, in these cases, learn more about our targets.

Of course, if we don't know anything about the nature of the target of our representations, it's hard to know in advance what will end up being 'noise' or extra structure, and so it might be difficult to determine what we should count as an RRT and what we should not. Note, though, that there is a sense in which this problem isn't unique to RRTs: there is a similar problem with regular old representation theorems. As Suppes and Zines point out, we can also construct “pathological” homomorphic functions from empirical systems to mathematical systems, that is, there are plenty of useless homomorphisms that hold in the case of regular representation of

empirically observable phenomena, at least when the target domain is finite or denumerable, and we need to somehow systematically exclude those from the useful ones, which we count as genuine representation theorems. And the suggestion about how to do so is not greatly elaborated on:

It is of value... to exhibit a numerical system that is not only isomorphic to an empirical system but employs certain simple and familiar relations as well. A complete or precise characterization of the intuitively desirable relations is unfortunately somewhat elusive, so for this reason the statement of the representation problem refers merely to an “appropriately chosen” numerical system (1963, p. 8).

So there is a similar problem with regular representation theorems: there are *too many* isomorphisms hanging around, and the onus is on us to determine which ones count as genuine. When it comes to representation theorems involving numerical systems, we might have stronger intuitions about which relations to favor in using those numerical systems to represent empirical targets (or, I suspect more likely, it just might be more obvious which relations are going to be *useful* to us). But similar problems to those I raised above complicate regular representation theorems as well.

I hope, in this section of the paper, to have done two things: first, to have simply fleshed out an account of radical representation and how it applies generally to cases where we lack empirical access to target phenomena; and second, to have given some motivation for thinking, once we have adopted the radical representation framework, that logical realists should think of soundness and completeness as together forming an RRT.

3. Further Discussion

I've argued that the logical realist should think of logic as *triadic*, that is, as consisting of a triad

of syntax, semantics, and the *target* that syntax and semantics are aiming to represent. I've argued that the logical realist should take the logical structure of the world to be that target. And I've given an account of how we should think about the relation between syntax and semantics in their role as representing the target, by giving an account of *radical representation theorems* and showing that we can take soundness and completeness together to form an RRT that shows us that if one representational side of logic succeeds at representing its target, so does the other.

This is of independent interest, especially to the logical realist, but there are also further questions that it raises. One of the most interesting of these is whether, generally, RRTs help *reveal* anything about the nature of their targets, and whether, particularly, the one under discussion here does. Recall that one of the things that distinguishes RRTs from their more familiar cousin, the regular old representation theorem, is that representation theorems relate an *empirical* structure to a numerical one, that is, they essentially relate an observable target (domain) and a numerical one. But we don't have a way to empirically observe the targets of RRTs. So how do we know what they are like?

I won't give a definitive answer to this question here, though I will suggest, eventually, that RRTs *do*, at least in some cases, teach us something about the nature of their targets. Most of this section of the paper, though, will be devoted to outlining what the possible answers to this question are, sometimes gesturing at how we might defend or develop them.

First, let's get two theses on the table. The first is that RRTs *give us* the nature of a target, or at least, that once we know of *all* the adequate representations of a given target, we can know what its nature is, roughly because *all there is to its nature* is the structural properties that its representations have in common. I'll call this the *strong thesis*.⁴⁴ The second is that RRTs *help*

⁴⁴ There is some relationship between the strong thesis (and more generally the account of representation developed here) and Wright's (1983) and Hale's (1987) neofregean views about abstraction, and there is also a more clear relationship between a kind of structuralist picture of, say, mathematics and the strong thesis. But

reveal the nature of their target by telling us what is *not* in the target: that is, that they tell us to ignore any features of our representations that aren't common to all of the representations that successfully represent a given target. I'll call this the *weak thesis*.

Let's start by thinking about regular representation theorems, by focusing on a case where we can empirically observe a phenomenon (but where there is widespread disagreement about the underlying metaphysics of that phenomenon!): mass. The strong thesis would tell us something like the following: what mass *really is* is just a bunch of structural relations; there is nothing more to mass in the world than these structural relations. (The strong thesis is going to immediately give us a pretty strong structuralism about whatever the target of representation is.) The weak thesis would tell us something like: we can learn some things from representation (and uniqueness) theorems when it comes to mass. For example, we learn that there is no metaphysically privileged *unit* of mass, or, more accurately, that none of the possible unit choices *we* can make when it comes to mass is metaphysically privileged. That is, that mass doesn't fundamentally come in grams, or kilograms, or shrams, etc. We learn this simply by noting that the various adequate representations of mass structure don't share a unit; and by applying the weak thesis, which says that anything not shared by all of our representations is not in the nature of the target.

One might wonder, given this example, what the difference between the strong and the weak thesis is. It will be more clear how they come apart when we consider abstract objects, but note, for now, that with respect to mass it is consistent with the weak thesis that there are, for example, infinitely many mass properties, each with its own quiddity, whereas this is inconsistent with the strong thesis.

there are too many important differences between this paper and these two views for me to adequately elaborate upon here.

What does rejecting both these theses amount to? It seems to me to commit us to the claim that one of our representations is *canonical*: that *it*, but not our other representations, somehow reflects the true nature of the target.⁴⁵ In the case of mass, this might amount to the claim that there *is* a privileged mass unit, and that the representation that uses that mass unit is the canonical one. It's also important to note that we should distinguish between the *epistemic* claim that we can't know which, if any, representation is canonical (which leaves open there being a metaphysically privileged mass unit), and the *metaphysical* claim that there *is no* canonical representation (because there is no metaphysically privileged mass unit)--though, as in a similar case in the first paper, there might be a relationship between the epistemic and metaphysical claim.

I'm going to refrain from discussing whether we should believe either the strong or the weak thesis when it comes to mass, because there are significant differences between this kind of case and cases of radical representation, where we can't make empirical observations about the target phenomenon.

What should we think about the strong and the weak thesis applied to radical representation? The first thing to note is that while the framework I am approaching the question from is somewhat unfamiliar, the general issue is not. Consider geometry: there is a close parallel between the question I am asking here and the question that Hilbert (1980) was considering when he suggested that we could replace lines and points with items like tables, chairs, and beer mugs: is geometry essentially about points and lines? Is it essentially even about relations in three dimensional space? Or is it just some kind of purely formal system? I suspect

⁴⁵ Being slightly more careful: one could reject the strong and weak theses and maintain only that one representation was *more* canonical than others, if one thought that being canonical comes in degrees. This is a funny way to talk given our normal use of 'canonical', but I am using it as a technical term here and I want to remain open to the idea that canonicalness is a gradable property. I won't say much about this here because the landscape looks similar regardless of which approach we take.

the answer is somewhere in between, but I won't take a stand on that here. Instead I want to point out that the extreme view—no holds barred structuralism about geometry—is motivated by the strong thesis. If we have RRTs that hold between purely algebraic formulas, pictorial representations, and so on, and we take our target to be *the equilateral triangle*, it's important to point out that *both* the strong and the weak theses seem to tell us to conclude something stronger than that geometry needn't be about points and lines. For while algebraic representations might be thought of as involving a kind of abstract, mathematical space, they don't seem to deeply invoke a notion of physical space in the same way that, say, a two-dimensional *drawing* of a triangle does. It follows that, according to both theses, geometry needn't be about *physical spatial* relations at all.

This is what structuralists about geometry want us to conclude. In the case of geometry, there are potential considerations that might weigh against this. First, we might think that what geometry is either *about* our concepts, or is about abstract objects that *correspond* to our concepts; and, our concepts of geometric relations are essentially physical-spatial. Second, we might think that geometry started off as *applied mathematics*, and that RRTs don't, and can't, change what the intended target of geometry is. These two considerations are related, and what they have in common is important: an appeal to what we *intend* the target of our representations to be. They both suggest that this intentions are enough to get us the claim that one of our representations is canonical.

One brief motivation for the claim that intentions are enough, when it comes to geometry, might be to see that there seems to be a contrast between this case and the case of mass. It is perhaps built into our very concept of a triangle is that it's a figure that lives in two-dimensional space, with three sides and three angles. But it does not seem to me that we have similarly built

into our concept of mass that there is a single privileged mass unit. I'm just speculating here, possibly based on purely personal anecdotal experience, but there was a time when I found the idea that geometry might not be about physical space at all surprising, but not so with the idea that mass properties might not come specially packaged in gram units. But how far should we go with these sorts of intuitions?

The question here is whether we have any reason to believe that the *way* we conceive of our representations as representing abstract objects and abstract structure maps onto the genuine nature of those objects. And if we are realists about those objects and that structure, there is going to be tension in claiming that the answer to this question is yes, unless we have an independently motivated view that allowed for this. We might worry, for instance, that it would be *arbitrary* to assume that a given representation was canonical without some independent account of the nature of abstract objects. And we might have more general worries about the epistemology of abstract objects. I am going to discuss two views that allow the realist to avoid these worries.

The first view I will consider is a kind of *plentitudinous platonism* (like that of Balaguer (1998)), which says that there are abstract objects that correspond to *all* of our possible concepts. On this view, we could maintain that (when it was abstract) the target of our representations was always the *intended* target of our representations, because we would always be *guaranteed* that there would be such a target. Hence, our intentions matter here, and help determine what object we are referring to; but the epistemic worry is arguably dissolved, because *all* of the possible objects we could intend to refer to are out there, and the way we can know what the nature of the one we are referring to is by picking it out via our intended canonical representation.

Is plentitudinous platonism a genuine alternative to the strong and the weak theses? No,

or at best, sort of. While plentitudinous platonism *does* resolve the question of how we could know that our canonical representations were canonical, it doesn't follow that RRTs, for the plentitudinous platonist don't *teach us* something about the nature of our targets. The plentitudinous platonist should adopt the following view when it comes to radical representation: RRTs don't show that two different representations *only* represent a single abstract object, because the plentitudinous platonist is committed to there being a distinct object for each representation (every representation of an abstract object is canonical). But they do show that *in addition* to the distinct abstract objects each represents, there is another object that they both successfully represent. This is because they *give us* a new concept: the concept that is equally well represented by one of the candidate representations or the other.

Consider a simple example of a sentence, “the candle is bent”. One way to represent this sentence is by writing it down. Another is by saying it out loud. The plentitudinous platonist might think that the written sentence is a *token* or representation of one abstract object which has some essential property that only can be represented by writing, and the said-out-loud sentence is a representation of another, which has some essential property that can only be represented by saying-out-loud. But if so, her view seems to dictate that she now needs to grant that there is a third abstract object which has neither of these essential properties: both the written sentence token and the said-out-loud sentence token represent some more general object, *the sentence* (perhaps also, or maybe instead, *the proposition*) “the candle is bent”. In general, the plentitudinous platonist should think is just what anyone else who believes in abstract objects should: that what RRTs show is that two representations both represent the same *target*. (Whether they *also* each represent a separate target is a different question.)

Where does this leave us? One reason we might want to reject the strong and the weak

theses is if we think that our *intentions* and *concepts* matter to whether a representation is canonical. But the realist faces a worry here: we have to be careful not to fall into a view on which our concepts and intentions *shape* reality. The worry is that it would be *arbitrary* to think that the abstract objects that exist are just those that we take our representations to aim at. The plenitudinous platonist has a solution to this arbitrariness worry: there is an abstract object for every concept we have, indeed every possible concept we could have. But I've suggested that there is an important sense in which the plenitudinous platonist shouldn't reject the weak thesis after all: while she should maintain that our intentions *can* matter in determining which of our representations are canonical, she should also grant that RRTs can demonstrate the existence of abstracta that none of our representations canonically represents.

A second view, which, due to its similarity in spirit to a view defended by Williamson (1994), I'll call *epistemicism*, avoids the arbitrariness worry raised above in a very different way. The epistemicist maintains that while we have no reason to think that any *particular* representation of a target is canonical, we can still maintain that *some representation or other* is canonical. She does so by maintaining that it's no barrier to one of our representations being canonical that we don't seem to be able to have epistemic access to which one is.⁴⁶ Unlike plenitudinous platonism, epistemicism *does*, I think, successfully completely avoid the strong and the weak theses, though one might think that this is a tradeoff because of other theoretical costs associated with the view.

There might be other views that support the existence of abstract objects that correspond to each of our concepts and don't fall prey to arbitrariness worries, but I won't consider any further here. I suspect that the strongest reasons to reject the strong and the weak thesis will

46 For readers of this entire dissertation, there is an obvious correlation here to the thesis called 'privileged' discussed in the first paper.

come not in *general* positions like plentitudinous platonism and epistemicism, but in *particular* cases. Perhaps, for example, treating certain representations as canonical allows us to make scientific or mathematical progress, and that might give us reasons, in particular cases, to think that a given representation is canonical. Whether this is the right conclusion to draw from a given representation being scientifically helpful in this way is too large of a topic for me to broach here.

I instead just want to suggest that there is some sort of methodological principle that says something like “endorse (at least) the weak thesis unless you have independent reason not to”. I don't take myself to have just argued that there are no good reasons to reject the strong or the weak thesis, but rather to have gestured at some reasons it might be difficult to claim that they are generally false.

There's an interesting (but not particularly surprising) parallel here to be drawn to the discussion in the first paper. There, I argued for the disjunction of *privileged* (the view that some small subset of our logical constants is metaphysically privileged, but that we can't know which) and *unfamiliar* (the view that *none* of our logical constants accurately captured worldly logical structure), and gave some reasons to think that *unfamiliar* was a preferable thesis. Here, I won't rehash this argument, nor will I give an equivalent one about the nature of abstracta more generally. But it's worth noting how the views map over to the discussion of RRTs here.

Privileged, as a thesis about how our words represent the world, maps onto *epistemicism* here: both theses essentially say that one (or some small subset) of our representations of a given target are canonical, but we can't know which. *Unfamiliar* is motivated by the idea that we should take an overabundance of equally good logical constants to be a guide to the claim that, whatever the *target* of those constants is, it doesn't structurally resemble *any* of them. Hence, it

seems to map roughly onto the weak thesis: we should, the claim is, take RRTs to be a guide to the claim that the target of those RRTs doesn't look anything like any of the (non-overlapping) structure of the representations.

I didn't, in the first paper, discuss any views in depth that would map onto the remaining options here. Some sort of view that rejected *weak non-redundancy* (the claim that there are no unexplained necessary connections between fundamental facts) might roughly map onto the plentitudinous platonist view I discussed above, though it would be complicated to work out the details and the connection is much more tenuous than in the other cases.⁴⁷ I also didn't discuss what the view that would naturally map onto the strong thesis would be. It would be something like a strong metaphysically structuralist claim about the nature of the logical structure of the world: that there is *nothing more to it* than whatever structural properties all of its adequate representations share.

While it is worth drawing these parallels, the argument from the first paper doesn't successfully generalize here (at least not immediately). We shouldn't jump to the conclusion that (at least) the weak thesis is true in the case of abstracta more generally. We haven't established the important claim that *there is no good reason or argument for* taking one or the other representation as canonical when it comes to abstracta more generally. And while I am generally sympathetic to the weak thesis, there is much more space for resistance to this kind of epistemic argument here than there is to the kind of argument I made in the first paper.

I now want to return to thinking specifically about logic, taking soundness and completeness to form an RRT. What would the strong and the weak thesis (and also their

⁴⁷ One problem with trying to endorse something like plentitudinous platonism in the case of logical structure is that it's important that we leave open the option, discussed in depth in the first paper, that we shouldn't be *quantifying over* logical structure, but still should be *realists* about it. It's very hard to see how the spirit of the plentitudinous platonist position could be transferred to this sort of nominalist-realist position.

rejection) look like for the logical realist? *Do* we have reasons, in the case of logic, to endorse either thesis?

It's actually very hard to answer this question. The strong thesis tells us that all there is to the nature of a target is whatever structure our representations share. But it's very hard to say much more about what that structure *is* in the case where we are treating syntax and semantics as representations of the logical structure of the world. One way of thinking about the relationship between syntax and semantics and the relationship between them is that they preserve equivalence classes of sentences; so we might think that we could somehow take the structure of those equivalence classes to be the structure in question. But to do so both moves a bit too quickly (I think) and it's also difficult to see what it even *means* to take the structure of those equivalence classes to be the structure in question.

The weak thesis merely tells us that there is nothing about worldly structure that is either syntax-like or semantics-like (presumably while also maintaining that whatever structure overlaps between them is shared by their target). It could be that there is some other *way* that worldly logical structure is, that is not captured by either side of representation, nor by the structural features they share.

It's hard to say more about what either of these theses amount to when it comes to logic, and we might take that as a mark against the view that either is true in this case. That might be right, but I doubt it is a conclusive mark. One reason to think that it is not conclusive is that it coheres with the claim I argued for earlier: logical realists ought to be committed to worldly logical structure being deeply unfamiliar in this exact way. A second reason to think that it is not conclusive is that it is *also* hard to see what would motivate the rejection of the strong and the weak theses here.

How might we reject these theses? First, we might claim that one representation was canonical, and that worldly logical structure is actually either more proof-theoretic in nature (perhaps it is *rule-like*), or that it is more semantic in nature. We also might endorse a weaker version of this—something like, being a canonical representation comes in degrees, and one of our representations gets things *more right* than the other, but we still need to abstract away from *some of* the structural differences. (I'll treat these two claims as one here even though they might come apart in different ways.) Second, we might try to make the plentitudinous platonism claim: both representations are canonical, but they represent distinct targets.

It is totally unclear how the first move would go. That being said, it is, I think, something we ought to explore. While there is a rich literature in the philosophy of logic about something that looks superficially like it is responding to the question of whether logic is *really* proof-theoretic in nature or semantic in nature, the “really” here *doesn't* typically mean *metaphysically really*. On my reading of (most of) this literature, no one is talking, at least not directly, about what *worldly* logical structure is like—what the world, independent of our mind and language, is like. The metaphysical issue is not typically under discussion. I can't possibly say much about this here: are there reasons for, arguments for, evidence for, the world being more rule-like in its logical structure or more semantic-like? Typically, inferentialism and proof-theoretic semantics is coupled with a kind of logical anti-realism, but I think the issue of what the world is really logically like is too underexplored for us to assume that the world couldn't be more like the proof-theoretic side of logic than the truth-theoretic side.

What about plentitudinous platonism? Remember that the plentitudinous platonist solves the arbitrariness worry by claiming that our intentions help determine exactly what the targets of our representations *are*. But anyone who holds that intentions matter in this way is going to have

some trouble when there's no agreement whatsoever about what our intentions *are*. Consider the disagreement between inferentialists about logic and those who think of the semantic side of logic as primary (more specifically, consider someone who wants to give an account of the nature of logical constants in terms of their proof-theoretic role vs. someone who wants to do so on the semantics side of logic). If our intentions determine which abstract structure we are targeting, and there are two distinct structures that correspond to the semantics side and the syntactic side of logic, then *regardless of whether logical realism is true*, these people are talking past each other, and are having a merely verbal disagreement. they aren't really disagreeing about anything at all, because they are talking about different kinds of entities and structure entirely. And it seems important, if logical realism is true, that this be a substantive, and not merely verbal, dispute.

We could preserve logical realism here by claiming that what these two parties are arguing about is what some *third* structure is like—whether it is like syntax or semantics. But if it is just like one or the other of them, then it's hard to see why it should count as a third structure, rather than just being identical to one of the two already under discussion. And, in the earlier discussion, I suggested that the plentitudinous platonist needs to grant that there are, in cases of RRTs, targets that satisfy the weak thesis: that have none of the representational features that one, but not both, of our representations have. It follows that she needs to grant that, even if syntax and semantics each represent a syntax-like structure and a semantics-like target respectively, they *also* represent a third target that is neither syntax-like nor semantics-like. But if this is the target under discussion, then *both* sides are wrong about the shared target of representation.

I've discussed various views one might have about what RRTs show generally, and what

they might show in the case of logic specifically. My own (tentative) view, which I won't defend here, is that something like the weak thesis is true in the case of logic. This is compatible with there being syntax-like worldly structures, and semantics-like worldly structures. But it suggests that, if there are such structures, they are less general, and perhaps less fundamental, than the target we are in fact *aiming* to represent, which is neither syntax-like nor semantic-like. That target doesn't have, as features, anything that corresponds to one of syntax and semantics but not to both of them.

But again, I won't defend this here. The central, weaker suggestion I want to make is that, if it's the case that we've got no *independent* grasp on what the logical structure of the world is like, then we've got no defeater for the methodological claim when it comes to evaluating candidates for being the “one true” logic. I argued in the first paper that we don't have a grasp on what logical structure is like along a certain dimension—that it doesn't resemble any of our logical constants. Whether it “looks” more syntax-like or semantics-like, though, is a slightly different question. More work would have to be done before we could answer that question. (Of course, if we are using a logic for some other purpose, or dealing with a specialized logic, this might not be the case. For example, we might think that we *do* have an independent grasp of the target of a deontic logic.)

It does seem that the logical realist has no *obvious* reason to think that either syntax or semantics canonically represents its target. It's easy to see that if we weren't logical realists, we could come up with such reasons. For example, if we took the target of our logic to be, as many do, our intuitive concept of validity or consequence or something like that, then one potential reason would be *what our intuitive concept of validity is like*. Think again about geometry: if geometry were about our intuitive concept of triangles and other geometric shapes, then we

would have a reason to think that pictorial representations of geometry were canonical: our intuitive concept of geometric figures is much more closely aligned with pictorial representations of triangles than with algebraic ones. Similarly, it might be the case that our intuitive concept of validity is much more closely aligned with either the semantic or the syntactic side of logic. (Though even here, there is no general consensus about whether our intuitive concept of validity is more like the syntactic side of logic or the semantic one.)

However, in neither geometry nor logic does this cohere well with realism about the domain they are targeting. If geometry is only about our intuitive concept of triangles and other geometric shapes, then it is hard to see how we can justify being realists about geometry. Likewise, if logic is only about our intuitive concept of validity, then it is hard to see how we can justify being realists about logic, in the sense under discussion in this dissertation.

A final, but important, note: the scope of positive argument in this paper is meant to be fairly minimal. An RRT doesn't tell us that a given syntax and semantics pair has, as its target, the world itself, or the logical structure of the world. What it tells us is that *if* we have an RRT for such a pair, then (a) that pair successfully represents the same target and (b) in the absence of defeaters, that target likely doesn't look like either member of the pair. The logical realist then needs to go on to *triangulate* between whatever information she has (or lacks!) about the logical structure of the world itself, and the syntax-semantics pair. If we have independent information about how the world is actually logically structured (perhaps from science, or from agreed upon norms of rationality, etc.--not taking a stand here about whether we have any such evidence, or even could have any such evidence), then we can check for how that information jibes with the *commonalities* between a given syntax-semantics pair.

4. Equivalent Theories, Definitional Extension, and Quantifier Variance

What is it for two theories to *say the same thing about the world*? When do two theories make genuinely distinct metaphysical commitments? While metaphysicians have often discussed these questions in some form or another, they have rarely posed them in the following way: what is it, generally, for two theories to be equivalent? More often, the question of equivalence, while it underpins debates about *verbal disputes* and the *substantivity* of certain metaphysical theses, is left lurking in the background. The aim of this paper is to instead tackle the equivalence question head-on.

One proposal about metaphysical equivalence, which comes to us from Glymour (2013) and de Bouvere (1965), is that in order for two theories, T and T', to be equivalent, they must have a *common definitional extension*. Intuitively glossed: if there is some way to *extend* T and T' into a third theory, T+, where any vocabulary in T+ but not T is definable purely in the language of T, and any vocabulary in T+ but not in T' is definable purely in the language of T'. I show that we ought to take this proposal seriously, by showing that two theories' having a common definitional extension (hereafter 'CDE') is an appealing *necessary* condition for equivalence. But, I argue, it is false. I argue that a related, but weaker condition, CDE*, is necessary for equivalence.⁴⁸

Since I'm only considering a necessary condition for equivalence, if it is too weak, the project here will be uninteresting. But CDE* is not weak at all. I will show that treating CDE* as necessary for equivalence entails that certain disputes in metaphysics *must* be substantive, since the competing positions in the disputes do not have a CDE*. I focus on composition: there is no

⁴⁸ Note that neither de Bouvere nor (arguably) Glymour are using common definitional extension for the same purpose that I am.

CDE* of nihilism (the view that composition never occurs) and universalism (the view that any two distinct objects compose an object). This will generalize to similar disputes in metaphysics. This is a powerful result. Treating CDE* as necessary for equivalence forces certain metaphysical disputes to be substantive *without* making any of the heavy-duty metaphysical commitments that are commonly taken on board in order to respond to the quantifier variantist.

So there are two related theses of the paper. First, that a necessary condition for two theories' being equivalent is that they have a CDE*. And second, that this rules out quantifier variance without any heavy-duty metaphysical machinery.

The paper proceeds as follows: in section 1, I lay out some background assumptions and further clarify the notion of equivalence at stake. In section 2, I introduce the notion of a CDE. In section 3, I show that CDE as necessary for equivalence has multiple virtues. In section 4, I compare CDE to another account of equivalence, and address a related objection. In section 5, I explain why CDE, despite its appeal, is problematic: it rules out paradigmatic cases of equivalent theories. I propose a new condition, CDE*, which avoids this problem. In section 6, I show that CDE* does real work for us by arguing that it rules out the possibility that nihilism and universalism about composition are equivalent.

1. Goals and Background Assumptions

I'll first locate the pre-theoretic notion I am trying to zero in on. We have some idea of what it is for two theories to be *empirically equivalent*. And while verificationists, for example, should say that that is just all there is to the kind of *metaphysical* equivalence that I am concerned with, anyone who thinks there are genuine distinctions between empirically equivalent possibilities must think that *metaphysical* equivalence requires more than empirical equivalence.

One might think that metaphysical equivalence is parasitic on an account of meaning—that if two theories mean different things, it follows that they do not say the same thing about the world. This isn't right. Metaphysical equivalence (I'll often drop 'metaphysical' in what follows) is not as fine-grained as *meaning equivalence*. There is some genuine distinction between the meaning of the sentence 'Kareem is taller than Annie' and the sentence 'Annie is shorter than Kareem' (whether that difference is a difference in propositions expressed is a more delicate question). But these two sentences are equivalent with respect to all of the worldly facts. They say the same thing about the world, and so they are metaphysically equivalent. So we can't settle the equivalence question by settling the meaning question⁴⁹

While a notion of equivalence lurks in the background of discussions about the meaningfulness of certain philosophical disputes, equivalence is more often explicitly discussed in the philosophy of science.⁵⁰ One might initially think that equivalence is a different matter in science than it is in metaphysics, and that equivalence of physical theories should be determined internally to physics. Perhaps, for example, there is a notion of equivalence-within-physics that is a matter of being equivalent with respect to everything we care about/are trying to accomplish/our particular scientific goals. But this is not metaphysical equivalence (unless the two happen to be coextensive).

I assume that there is some objective, goal-independent fact of the matter about whether two theories say the same thing about the world. And such an account of equivalence should be generalizable, since it is hard to see how it could vary across domains. There is hence no

49 For a similar point raised from an entirely different direction: Miller (ms) argues in depth for disentangling two notions of hyperintensionality—she might not put it this way, but “in the head” vs. “in the world” sources of hyperintensionality, where only the latter tracks genuine distinctions in the world. (So the 'believes that' operator typically gets us “in the head” hyperintensionality, but, e.g., perhaps 'grounds' tracks worldly hyperintensionality.) We can see that if this is right, and I think it is, meaning-equivalence, if it tracks *both* kinds of hyperintensionality, can't be a guide to metaphysical equivalence.

50 e.g. North (2009), Barrett (2015), Glymour (1970, 1977, 2013), Halvorson (2013), Barrett and Halvorson (forthcoming).

meaningful distinction to be made between scientific theories and philosophical theories with respect to the target notion.

One could deny that there is such a notion of equivalence, but doing so comes at a cost: it pushes one fairly far into anti-realism about both science and metaphysics. Part of my goal is to develop an account of equivalence that allows for us to maintain a kind of scientific and metaphysical realism—to commit to the idea that the world is such that some empirically adequate descriptions of it are, metaphysically speaking, better or worse than others. But I also hope that this account doesn't commit us to endorsing metaphysical notions that some take to be esoteric. I attempt to find some middle ground between, on the one hand, the “anti-metaphysicalist” (who thinks that the world is a structureless blob), and, on the other, the “metaphysicalist” (who deploys notions that many of us are suspicious of in order to respond to the anti-metaphysicalist).⁵¹ I offer a route towards thinking that the world is structured, and in fact, is highly, objectively structured, that one can endorse without endorsing fundamentality talk, grounding talk, Lewisian naturalness, or any associated notions.

I focus entirely on first-order theories. This is a limitation of the project, but it is one that is hard to avoid. It is not that we *ought* to focus entirely on first-order theories. It is just that giving accounts of equivalence is extremely hard in even the simplest cases. So I focus on a range of (fairly) simple cases, acknowledging that we would need to think carefully about how to extend (and whether we could extend) the spirit of what I say here to other cases.

I also want to say something about my goal in the paper: to give a *methodologically useful* account of equivalence. I am looking for criteria for equivalence that can actually guide us in determining, in particular cases, whether two first-order metaphysical or scientific theories are

⁵¹ I borrow these terms from Rayo (2013).

equivalent. This will come in the form of a necessary condition for equivalence, so we can think of it as a *test* for *inequivalence*.

This goal might conflict, in certain ways, with the goal of giving an *analysis* or *real definition* of equivalence, and of determining whether equivalence is reducible to other concepts. To see why, consider Kristie Miller's (2005a, 2005b) account of equivalence in terms of truthmakers: that the question of theoretical equivalence is at bottom a question about *when two theories have the very same truthmakers*. Note that this kind of account doesn't *help us* decide on particular cases of equivalence.

Claiming that two theories are equivalent just in case they have the same truthmakers doesn't give us any methodological guidance, for it doesn't tell us anything about how to figure out whether two theories have the same truthmakers. It might be right that equivalence admits of a *analysis* in terms of sharing truthmakers. But in terms of methodological progress, this merely pushes the bump around under the rug: we've got no more access to when to theories share the same truthmakers than we do when they are equivalent in the first place. This may be because the notion of sharing the same truthmakers is simply too *close to* the notion of equivalence for the analysis to be helpful. In section four, I'll show that the same “methodological guidance” problem that faces truthmaker equivalence arises for other accounts of equivalence, but I should note that this perhaps shouldn't be construed as a *problem* for these accounts, since their aims are simply different than mine.

2. What is CDE?

First, let me get on the table what it is for a theory, T^+ , to be a definitional extension of a theory T .

T^+ is a *definitional extension* of T (where T^+ is a theory in the first order language L^+ , T is a theory in the first order language L , and L is a subset of L^+) iff the following conditions hold:

First, T^+ is a conservative extension of T . (This just means that every theorem of T is a theorem of T^+ , and that every L -theorem of T^+ is a theorem of T .)

Second:

a) for every relation symbol of L^+ there's a formula ϕ of L such that:

$T^+ \models \forall x (Rx \leftrightarrow \phi(x))$

b) for every function symbol f of L^+ there's a formula ϕ of L such that:

$T^+ \models \forall x, y (f(x) = y \leftrightarrow \phi(x, y))$

c) for every constant symbol c of L^+ there's a formula ϕ of L such that:

$T^+ \models \forall x ((x=c) \leftrightarrow \phi(x))$

(Note that in each of these cases, the L -formula need not “match” the L^+ formula—so e.g. in a) $\phi(x)$ might be non-atomic.)

This amounts to saying something like the following: we can extensionally “define” every relation symbol, function symbol, and constant in T^+ using a formula of L . For shared symbols of L and L^+ this will be trivial, but for non-shared symbols it is not. So there is certainly some sense in which we can “build up” T^+ out of T .

Two theories, T (in L) and T' (in L') have a *common definitional extension* (hereafter 'CDE') T^+ (in L^+) iff T^+ is a definitional extension of T and T^+ is a definitional extension of T' . What we want is for T and T' to say all, and only, the same things. But this makes it sound like they have to express the very same *propositions*. I don't think this is the case (see my comments in the introduction about theoretical equivalence being less fine-grained than meaning equivalence). What CDE gets us is the following: assuming that T and T' are complete, for any set of objects that T singles out and says a true thing about, T' also singles out that set of objects and says a corresponding true thing about it.

de Bouvere (1965) proved that, if T has a definitional extension, T^+ , then for every

model M^+ of T^+ , there is a unique model M of T that is the result of removing any additional L^+ structure (that isn't L -structure) from M^+ . Unless the extra structure that L^+ has that L lacks represents a genuine feature, division, or “joint” in the world, this seems to mean that M^+ and M are modeling the same possibility.

An example may help here. Suppose that T is a theory in a language that contains the predicates 'green' and 'blue', as well as sufficient temporal language to extensionally define 'grue' and 'bleen'. Let T^+ be the definitional extension of T which is in the language L^+ , which adds only two terms to L , 'grue' and 'bleen'. Then a model, M^+ , of T^+ , will involve a specification of the extension of 'grue' and 'bleen'. A *forgetful mapping* from L^+ structures to L structures will just *erase* the specification of the extension of 'grue' and 'bleen' from the L^+ structures, turning the class of L^+ structures into a class of L structures. The resulting L structures will be exactly those L structures that are models of T . This mapping is 1:1. Further, the members of the pair of models $\langle M, M^+ \rangle$ will have the same number of elements in their domains. (We could instead delete 'green' and 'blue' from T^+ , turning the class of L^+ structures into a class of L' structures, where L' is the language that contains 'grue' and 'bleen' but not 'green' and 'blue'.)⁵²

Why are these model-theoretic results important? First, they help demonstrate why many metaphysicians will want to reject CDE as a *sufficient* condition for metaphysical equivalence. I haven't shown that the grue/bleen theory and the green/blue theory have a CDE, but I think it's easy to see that, given enough linguistic resources in each theory, they do. Many of us will want to deny that these theories are *equivalent*, because we think 'green' and 'blue' are *more metaphysically privileged than* 'grue' and 'bleen'. If we want to maintain any distinctions in this ballpark, we have to reject CDE as a sufficient condition for equivalence.

⁵² These results are shown in de Bouvere (1965).

Second, CDE is, on its face, an entirely syntactic criterion for equivalence. I am working with some sort of view which treats theories as *sets of sentences*, rather than a view that individuates theories by their models. One might object to this. I follow Halvorson (2013) and Lutz (2015) in thinking that the differences between a (plausible) semantic view of theories and a syntactic view of theories are illusory. And given my methodological goals, the syntactic approach works best here. But there is something else to say here: as de Bouvere (1965) points out, and Glymour (2013) emphasizes, two theories' having a CDE *entails* that their classes of models *also* bear this intimate relationship to one another. That CDE gives us a connection between our models and our syntax may help mitigate worries that arise if we do maintain the distinction between the semantic and syntactic views of theories. I'll (briefly) discuss this more in section four, but what is generally important here is that a CDE doesn't just preserve linguistic structure in our theories—it also preserves, via our “forgetful mapping”, model-theoretic structure.

CDE is not quite what we want, for reasons that I will elaborate on in section 5 of the paper. There, I will suggest that we need to tweak CDE (to get “CDE*”) so that we must also be able to define up, in T and T' , any *logical constants*—quantifiers, connectives, and so on—that appear in T^+ . Otherwise, CDE fails to capture some fairly trivial equivalences between theories. But the central *motivations* for adopting CDE* are no different than those for adopting CDE, and so first, in sections 3 and 4, I discuss motivations for, and objections to, CDE.

3. Virtues of CDE as Necessary for Equivalence

I'll offer two motivations for treating CDE as necessary for equivalence. The first provides us with a useful framework for thinking about the relationship between various views of

equivalence, which I will return to throughout the paper. CDE as necessary for equivalence fleshes out an intuition that many will share: that if two theories are equivalent, there must be some kind of possible perspective from which, when the theories are stated and contemplated, they turn out to be a single, unified theory. If two theories make the same metaphysical commitments, there has got to be *some* perspective from which they *look like* a single theory. If there weren't such a perspective, not even God could look at two theories and 'see' that they were equivalent. And in order for her to be able to know that when she restated the theories from her perspective, she was genuinely restating them, her linguistic framework would have to *include* the linguistic frameworks from which the theories were initially stated, or at the least, would have to allow her the resources to *construct* those linguistic frameworks.

CDE requires this, but is otherwise silent on what this perspective must be. The only requirement, which I'll say more about in a moment, is that it takes all of the language and structure of the two theories seriously.

I won't consider the view that there need be *no* perspective from which two theories look identical in order for them to be equivalent, partly because there can be no account of equivalence on such a view. But there are accounts of equivalence that seem to agree with the idea that there must be some such perspective, but make distinct claims about *what* perspective things must look equivalent from. Earlier, I mentioned a view on which metaphysical equivalence just is empirical equivalence. There are problems with clarifying empirical equivalence, but one can see that at least roughly, such a view is committed to the idea that there is a *particular* perspective from which two theories must look identical, and that that perspective just is *our* perspective qua observers of the world.

Fundamental equivalence says that two theories are equivalent just in case they are

fundamentally equivalent: just in case they have the very same *fundamental* commitments. I'll discuss a specific version of this view, (tentatively) defended by Sider (2015) in the following section. For now, I just want to motivate the idea that it is, like empirical equivalence, committed to the idea that there must be a *particular* perspective from which two theories need to look identical for them to be equivalent. What is this perspective? On one metaphorical way of characterizing fundamentality, the fundamental is the core of reality that “God would have to make to get everything else for free”. So fundamental equivalence suggests that it is something like *God's* perspective from which two theories need to look identical in order to look equivalent.

Since I'm only extolling the virtues of CDE as a necessary condition for equivalence, so far it doesn't look like it conflicts with either of these views. Later, I'll explain why this isn't quite right, but for now, I just want to get this intuitive motivation—the idea that all CDE is doing is requiring that there be *some perspective from which* two theories look identical—on the table.

The second virtue of CDE is that it nicely accounts for an intuitive test for merely verbal disputes proposed by Chalmers (2011). Chalmers' test is this: first, identify the term(s) in the debate that are potential sources of a verbal dispute (so if I claim “cars drive on roads” and you claim “cars don't drive on roads! Cars swim in the ocean!”, for example, equivocation over the term 'car' is a potential source of our debate's being merely verbal). Second, *eliminate* the offending terms and try to find a new sentence that we disagree over the truth value of. (In the elimination process, we are allowed to introduce new terms into our joint language, e.g. 'car₁' and 'car₂'.) If we can find one, then the dispute is substantive, whereas if there is no such sentence, then it is merely verbal (since we have identified as the sole source of disagreement the

way in which we were each using the term 'car'). (Chalmers 2011, p. 11.)⁵³

Chalmers' test squares well with treating CDE as necessary for equivalence. If the above example involving 'car' is a genuine example of a verbal dispute, then eliminating the offending term will result in two new terms (in this case, 'car₁' and 'car₂'), the meaning of which will be agreed upon and understood by each party in the purported disagreement. We can now agree that 'car₁s drive on roads' and that 'car₂s swim in the ocean'. Nothing is left that we disagree about. Hence, Chalmers claims, our dispute was merely verbal. we weren't initially saying distinct things about the world, we were just using different languages to express the same truth about the world, in the same way that 'snow is white' and 'la neige est blanche' express the same truth about the world.

Our new shared language, which includes both 'car₁' and 'car₂', will be a CDE of my original language and your original language. Every new predicate (in this case, only 'car₁' and 'car₂') will be definable in your original language and in my original language. And the new language will be a conservative extension of each of our original languages. Insofar as Chalmers' methodological criterion is appealing, CDE ought to be as well.

There is one caveat: we can't construct a CDE of our two theories until *after* we disambiguate. (Before doing this, there is a sentence in our common language which we disagree about the truth value of, and we haven't barred the term that is the potential source of verbal dispute.) But this isn't a great cost: we can be liberal about potential sources of equivocation, introducing new terms whenever we are suspicious. If we are using a term the same way, then when we attempt to construct a CDE of our theories, our two replacement terms will end up having the same definition, and we will be able to construct a sentence of our new language

⁵³ There's an exception here for cases of 'vocabulary exhaustion'. So it might be the case that sometimes a disagreement is genuine, but we lack the words to be able to express a new sentence, after the elimination process, that we disagree over the truth value of.

which we disagree over the truth value of. So, in such cases, we can *show* (by showing that our two theories lack a CDE) that we weren't having a merely verbal dispute after all.

So, insofar as one thinks that Chalmers' test for verbal disputes is a good one, one should think CDE is a good necessary condition for equivalence.

4. Objections and Alternatives

Let's return to our “perspective” talk for a moment. Recall that CDE requires that there be *some* perspective from which two theories look equivalent, and that fundamental equivalence requires that it be something like *God's* perspective from which two theories look equivalent. This makes it sound as though treating CDE as a necessary, but not sufficient, condition for equivalence is *compatible with* both fundamental equivalence. But this isn't right.

CDE requires our perspective, whatever it is, to preserve *all of* the linguistic structure in each theory: we must be able to define (in both T and T') all of the new vocabulary in T+.

Fundamental equivalence rejects that picture, and allows us to *ignore* some of the linguistic structure in our theories. The proponent of fundamental equivalence thinks we ought to be able to ignore non-fundamental commitments in our theories, since she is pre-committed to the idea that any non-fundamental commitments bottom out in fundamental ones. It is only the fundamental commitments that metaphysically matter. In what follows, I'll clarify what this means, and argue that it is a bad-making feature of fundamental equivalence *if* we want our account of equivalence to do philosophical work for us.

I'll then turn to a seemingly distinct challenge to CDE: that it assumes a problematically syntactic view of theories. I'll show that, in fact, attempts to elaborate a view of equivalence that assume a strong semantic view of theories fall prey to the same kind of problem as fundamental

equivalence.

Fundamental Equivalence

On one way of characterizing fundamentality, the fundamental is the core of reality that “gets us everything else for free”. If this is right, we should think that any non-fundamental difference is *grounded in*, can be *defined* in terms of, or at the least *supervenes on* a fundamental one. And this suggests that realists who buy into fundamentality and associated notions could give an account of equivalence in terms of fundamental equivalence. This is Sider's view: two theories are metaphysically equivalent just in case they bottom out in the same *metaphysical semantics*.⁵⁴

On this view, we *take away* structure from two theories to get a third theory from which those two theories look identical. In fact, structure minimization is required for evaluating non-fundamental theories for equivalence: we evaluate theories for equivalence by ignoring all of their non-fundamental structure, solely focusing on their fundamental commitments.

Do fundamentally equivalent theories always have CDEs? This depends on two choice points: first, do theories always *contain* their fundamental cores? And second, can always explicitly define a theory's non-fundamental vocabulary in terms of fundamental vocabulary?

For Sider, theories clearly *don't* need to contain their fundamental cores. If a nonfundamental theory has a perfectly fundamental *metaphysical semantics* (essentially, if it can be defined purely in fundamental terms), then it is true if the resulting fundamental theory is true. But it needn't be the case that those fundamental terms are *explicitly contained within* the theory. Sider's answer to the second question is yes: if a theory is *true*, this is because its metaphysical semantics will provide exactly the definitions in question. All true theories can be

54 This comes out in Sider (2011), but is articulated much more explicitly in (2015).

defined using fundamental vocabulary.

It follows that *if* two (true) theories had fundamental cores that were entirely *explicitly* contained within them, *then* those two theories being fundamentally equivalent would entail that they had a CDE. Hence, there is a sense in which the two views are compatible. But before evaluating whether two theories have a CDE, we would have to put them in “canonical fundamental form”: make all of their fundamental commitments explicit. If we could do so, then if they had the same fundamental commitments, they would have a CDE: all of their respective non-fundamental commitments would already be definable in terms of their fundamental commitments, so if they shared a fundamental core, they would have a CDE.

Fundamental equivalence suffers from an epistemic problem: there are many cases in which we can't know exactly what two theories fundamentally bottom out in—what it is about fundamental reality that makes them true—but in which we seem to know that they are equivalent. For example, we (arguably) know that the kilogram theory of mass and the shmilogram theory of mass are equivalent (a shmilogram = 1.2 kilograms) without knowing what the right fundamental metaphysics of quantities is. But it is a consequence of fundamental equivalence that we can't have such knowledge without knowledge of the relevant fundamental facts.

If we we don't know what a theory's fundamental commitments are, then we don't know how to put that theory into canonical fundamental form. CDE is a useful necessary condition for equivalence because it allows us to actually evaluate any two (first order) theories for equivalence. Fundamental equivalence, on its own, is *not* useful, since it says that two theories are equivalent just in case they share the same fundamental commitments, and we are (often) ignorant of this. We might be able to *reconcile* fundamental equivalence and CDE if we could

first put theories into their canonical fundamental form. But, I claim, we are usually too ignorant of fundamental facts to do so.

I won't take a stand on whether fundamental equivalence is *true*. Recall the claim that it is *God's* perspective, or perhaps *fundamental reality's* perspective, that two theories must look identical from for them to be equivalent. We have many reasons to think that this is right. But it's easy to see what the problem is: we've got no way to see things the way God does, or the way “the world” does. If anyone can access the facts about equivalence, though, it's God! Equivalence and fundamentality might simply be too *close* to one another for such an analysis to be useful to us—it is perhaps too tight of an explanatory circle to cash equivalence out in terms of fundamentality.

That being said, I emphasize that nothing about CDE as a necessary condition for equivalence entails that there is even any *sense* to be made of fundamentality talk, or that we must embrace it. I've shown is that there is a strategy for reconciling CDE with fundamental equivalence, but not that we must do so; and that there is an epistemic problem with fundamental equivalence which makes CDE a better *test* for inequivalence.

One thing that this reconciliation speech shows us is that *if* we are to be able to turn fundamental equivalence into something useful—something that guides us in evaluating theories for equivalence—we will first need to restate our theories in such a way that will end up forcing us to respect all of the *resulting* linguistic structure. And this pushes the epistemic problem to the level of what the right way to restate those theories *is*: how do we know what structure is fundamentally relevant, and what is not? Which structure can we chop off, because it doesn't correspond to a worldly difference?

If we knew what structure represented real differences in the world and what structure

didn't, we could chop of structure all over the place. And if we had such knowledge, we might want to reject CDE as too restrictive. Why force us to put theories into canonical fundamental form if we already know what their fundamental commitments are, and can go straight to those commitments, ignoring non-fundamental ones? If we were engaged in the project of giving an account of equivalence for the omniscient, we could certainly do so. But we mere mortals must be careful about structure-chopping.

One might object: why think that we need to know *what* the fundamental cores of our theories are in order to know that they have the *same* fundamental core? Perhaps we know that the kilogram theory and the shmilogram theory of mass have the same fundamental core without knowing what the correct fundamental theory of mass is. This is exactly right, but it is not a defense of fundamental equivalence as methodologically useful. Suppose we know that the kilogram theory and the shmilogram theory are fundamentally equivalent, but don't know what they are both fundamentally committed to. Then it can't be that we have this bit of knowledge—that *they are fundamentally equivalent*—in virtue of knowing that they make the same fundamental commitments. We don't know anything about what those commitments are! So we must have come to know this bit of information via some other route, not via examining their fundamental cores.

There is a related advantage of CDE over fundamental equivalence. One part of the motivation for fundamental equivalence is to block *quantifier variance*, the view that there are multiple, equally joint-carving quantifiers. Both Hirsch and Sider agree that quantifier variance must be true if it is the case that the universalist and the nihilist about composition are talking past each other. For that to be the case, there must be some term in their language they are equivocating over, and there are strong arguments to the effect that the existential quantifier is

the only candidate for being that term.

Hirsch's views, when unpacked carefully, should be unsettling to the metaphysical realist.⁵⁵ But Sider's (2011) account of fundamental metaphysics, partly intended to defend against quantifier variance, is unsettling for other reasons. Sider argues for a picture on which there is a single, ideal, perfectly fundamental theory of the world, which is stated in a language, the terms of which each carve nature perfectly at its joints (either by referring to some bit of fundamental ontology, or by perfectly mirroring, qua bit of “ideology”, the way the world is structured). One consequence of this view is that we are forced to think that every theoretical choice, even those that seem purely conventional, *matters metaphysically* when we are talking about the fundamental theory. So, for example, it turns out to be a hard, (and metaphysically meaningful) choice whether to use the existential quantifier or the universal quantifier in stating our fundamental theory. (See Sider 2011 ch 10.)

In section 6, I argue that treating CDE as necessary for equivalence has the virtue of making it the case that universalism and nihilism, for example (but also any other metaphysical dispute which seems to bottom out in whether quantifier variance is true), cannot be equivalent. But it doesn't, at least once we doctor it a bit (which I will do in section 5), have the unsettling consequences of Sider's view—it doesn't commit us, for example, to thinking there is a real difference between the existential quantifier theory of the world and the universal quantifier theory of the world.

I'll next turn to a seemingly different challenge: why assume such a syntactic picture of theories? Perhaps we ought to think about theories semantically, which suggests that we should give an account of equivalence by thinking about the *models* of our theories rather than the

⁵⁵ I should note here that Hirsch (2005, 2009) conceives of *himself* as a metaphysical realist, but that's somewhat inconsistent with the way I conceive of metaphysical realism. Hawthorne (2009) argues that Hirsch's views can only be motivated by a kind of verificationism, which I agree with.

language of our theories. I'll argue, however, that semantic accounts of equivalence suffer from the same problem that fundamental equivalence does.

Why Go Linguistic?

Why should we think that the perspective from which two theories look identical needs to be a linguistic perspective? Perhaps what matters, for example, is that there is a way to “see” the mathematical models of two theories as identical, and perhaps there is a way to do this non-linguistically, or at least, not in the sort of language that would be an extension of either of the languages that the initial theories are stated in. (Perhaps the theories are stated in English and English', for example, whereas there is some purely mathematical or spatially representational way of viewing the theories as identical. Perhaps God can simply look at the worlds that T is true at, and look at the worlds that T' is true at, and see the very same worlds.) Someone who holds the semantic view of theories might be pushed in such a direction—since theories are individuated by their models, equivalence between theories needs to be a matter of some sort of relationship between models.

I believe some version of the syntactic view of theories is right, or, more accurately, that insofar as we can make sense of semantic views, they end up collapsing into syntactic views.⁵⁶ But I will not argue for that here. Instead, I will focus solely on the epistemic issues with the semantic approach.⁵⁷

⁵⁶ For discussion of this issue, see Halvorson 2012, 2013, Glymour 2013, Lutz 2015.

⁵⁷ For those interested in criterion of equivalence the proponent of the semantic view could even give, see Halvorson 2012, 2013, Glymour 2013. To briefly summarize: One thing she might claim is that for two theories to be equivalent, their classes of models must be identical. Halvorson (2012) clearly shows that this is an unacceptably strong necessary condition for equivalence, because there may be trivial structural differences between models that we would all agree do not track potential worldly distinctions. Perhaps instead we could think of the perspective from which two theories look equivalent, for the proponent of the semantic view, as an *isomorphism*. There's some sense in which we might be able to understand two theories as identical from *within*, so to speak, an isomorphism that they bear to one another. The agent-centered language here is just to help see what I am getting at. The point is just that one way to think about what isomorphisms are is as functions which

Any *purely* semantic view of equivalence will be methodologically unhelpful, even if we grant that the semantic view is *true*. And this problem bears some similarity to the problem with fundamentality equivalence, though there are also important differences. What we get, from scientists and philosophers, is usually written words (and numbers) on a page. The challenge that is interesting is how we know when two purportedly distinct views actually say the same thing about the world. But we don't communicate our views to one another via presenting their models—divorced from a linguistic framework—to one another. Rather, when we communicate our theories to one another, we communicate linguistically. So any methodologically useful account of equivalence is going to be *routed through* language, even if some moderate semantic picture of theories is right. As Glymour says:

On the semantic view, to present a theory is to specify a class of relational structures. That can only be done through a description, in a language, of the class of structures... Other writers (Van Fraassen 1989, Lloyd 1994) ... suggest that model classes are directly given without the use of a language. What they, unlike Suppes, say cannot be taken entirely seriously because it yields a conception of theories that make them ineffable and their characterization magical. Newton, Einstein, Schrodinger, and so on, had no way of specifying the class of relational structures they intended, except indirectly, as those structures satisfying their theoretical claims. Nor do we today. The “yoke” to language may be neglected, but it remains... The space of models of, say, Newtonian theory cannot be

“see” two sets or classes as identical. So, we might think of particular defined isomorphisms as *themselves* perspectives from which two theories are identical, and perhaps we can think of them a *structural* perspectives rather than *linguistic* perspectives. Indeed, this is, I think, what the structural realists want us to believe.

This is a plausible way—though a somewhat metaphorical way—to think of what isomorphisms *are*. But I believe, following Halvorson (2012), that it is not plausible that isomorphisms between models are candidates for being the right *kind* of perspective to serve as a genuine necessary condition for equivalence. Saying that two classes of models are isomorphic tells us nothing until we specify the notion of isomorphism we are talking about. And it turns out none of the obvious candidate relations will work. As Halvorson points out, identity is far too strong to serve as a necessary condition for equivalence (and is a trivial sufficient condition); equinumerosity is an isomorphism, but it is far too weak to serve even as a useful necessary condition. (Theories that have *wildly* different subject matters may have equinumerous classes of models.) It's all well and good to claim that equivalence is a matter of structural equivalence, but unless we can make the notion of structural equivalence precise, we are left with no more information than that equivalence has something to do with the relationship between the models of theories.

One might react to this by pointing out the results in model theory, which I outlined in section 2, that we get in the case of CDE. Indeed, Glymour (2013) does so in response to Halvorson. Our “forgetful mapping” might be thought *itself* to be a perspective from which T and T' look identical. This is correct. However, as Halvorson (2013) points out, we can only get de Bouvere's results once we *buy into* a view of theories that doesn't just involve non-linguistic models; our models here *are* themselves linguistic structures, and we can only make sense of the de Bouvere results if we already accept that theories are linguistic entities.

directly indicated by pointing or looking in a magic closet or through a magical looking glass. Muggle science has no Diagon Alley (2013 p. 287).

The upshot here is that even if some version of the semantic view were right, we wouldn't have resolved the interesting problem, because our theories would, in practice, still be expressed in languages, and their models would still also be tied to those languages. There isn't some magical way to “see” that two purportedly distinct theories have the same models, without going through the languages of the theories themselves. We can't abstract away from the linguistic structure of our models to see when they are equivalent. So, a very strong semantic picture, e.g. Van Fraassen or Lloyd's, which individuates theories by their *idealized, language-independent* models, individuates theories on the basis of something that we don't have access to without first having done the sort of work I am proposing here. But a more moderate semantic picture, e.g. Suppes', which individuates theories by their models while recognizing that those models involve, and are deeply tied to, language, will still have to answer the question of how we evaluate theories in different languages for equivalence. And the model-theoretic results of CDE that I discussed in section 2 suggest that CDE is a promising condition for equivalence for the proponent of this kind of more moderate semantic view of theories.

There is a clear similarity between this issue and the problem with fundamentality equivalence. In both cases, we might think the accounts get something right—but they can't be good methodological guides, because our only route to evaluating *when* two sentences have the same fundamental commitments, truthmakers, or models, is via their linguistic commitments. It gets epistemic order of discovery the wrong way around, even if it gets things *metaphysically* the right way around.

Is there some perspective from which theories could look identical that is *neither*

linguistic nor about the relationships between models? I suppose it is open to someone who doesn't like the idea that such a perspective must be a linguistic one to propose some alternative. However, I take it that the burden here is on the proponent of such a view. And I suspect any non-linguistic view would fall prey to the same sort of problems the semantic view does. If the alternative approach attempts to divorce itself from language completely, it will face the epistemic problem, and the new challenge will be to show how we can know about the relationship between our language and whatever the object of the new approach is. Facing this challenge will force us back into thinking about the relationship between sets of sentences.

5. The case for CDE*

CDE has many virtues as a necessary condition for equivalence. But there is a problem: it rules out some trivial cases of theoretical equivalence. A background condition built into what it is for two theories to have a CDE is that the logical constants are held fixed between the two theories: that every logical constant in L is in L' , and vice-versa. The results in the model theory presuppose that this condition is met. But we want to be able to evaluate theories that contain different logical constants for equivalence. For example, we can't evaluate whether a theory that makes use of disjunction and negation is equivalent to one that makes use of conjunction and negation, even if the two theories seem to us to be mere notational variants. The dialectic here is a bit tricky. First, I'll clarify why I claim that this background condition is in place. Then I'll show how we can eliminate the background condition, and replace it with something more apt for evaluating theories for equivalence, by adding extra (explicit) conditions to CDE.

To motivate the claim that this is presupposed, let's think about what would happen if we didn't hold the logical constants fixed. Definitional extension says nothing about defining *logical*

constants in T^+ that don't appear in T , in terms of the language L . It only says things about 1) function symbols, 2) predicates, and 3) individual constants. We might think that the fact that T^+ must be a conservative extension of T will serve the purpose of forcing us to hold the logical constants fixed, and so it would be redundant to insert a requirement that we need to somehow be able to define the logical constants unique to L^+ in terms of L . At first this seems right: if T^+ is a conservative extension of T , then any logical constant introduced in T^+ will need to be conservative with respect to the logical constants that appear in T . And so maybe, the thought goes, it is not that we are *presupposing* that we have to hold our logical constants fixed, but rather that the conservative extension requirement *ensures* that we hold our logical constants fixed.

But we don't retain the spirit of what it is to be a definitional extension if we allow the addition of logical operators into T^+ that are conservative with respect to T —nor do we retain the model theoretic results I discussed in section 2. Suppose we start out with some non-modal first order logic, and conservatively add a modal operator to get our new theory, T^+ . T^+ is a conservative extension of T , and we've not violated any of the three additional conditions for definitional extension, so T^+ is a definitional extension of T . But we won't get the right results in model theory, since the models of T^+ will bear none of the resemblances to the models of T that we need them to (and there will be no way to give a forgetful mapping from the models of T^+ to the models of T). So the proof of the model-theoretic results for CDE will not go through unless we *assume* that we are holding the logical constants fixed, in addition to satisfying the explicit requirements.

If CDE is going to do non-trivial work for us (especially before we consider whether it is a sufficient condition), we need to rule cases like this out.

For consider again one of the motivations for CDE being necessary for equivalence: if two theories are equivalent, there must be some third “perspective” from which they look identical. But now suppose that $T' = T+ =$ a modal conservative extension of T , where T doesn't contain any modal operators. If we don't rule this out, either explicitly or implicitly, it's now the case that a theory and its modal conservative extension have a CDE, without it being the case that there is any third perspective from which they look identical (since certainly T and T' don't look identical from the perspective of T' , and T' is not anywhere near expressible using the resources of the language of T). So our intuitive motivation is shot.

The presupposition that we have to hold our logic fixed when evaluating theories for a CDE *also* rules out what I take to be the most obvious cases of equivalence, for example: suppose that the only logical constants in L are ' $\vee$ ', ' $\sim$ ', and the existential quantifier. And suppose that the only logical constants in L' are ' $\&$ ', ' $\sim$ ', and the universal quantifier. Imagine that every single other feature of L and L' are common to the two languages. Certainly, it seems, there could be a theory, T , and a theory, T' , which are equivalent. But the presupposition prevents us from simply expanding L to include ' $\&$ ' (or, equivalently, expanding L' to include ' $\vee$ ') to get a theory $T+$ in a language $L+$ that is a CDE of T and T' . $T+$ won't be a definitional extension of *either* T or T' if we are prevented from adding logical constants when producing a definitional extension.

There is a non-ad-hoc way to fix this problem: drop the implicit assumption, and add *explicit* conditions to what it takes to be a definitional extension that require new logical constants to be “defined” in similar ways to function symbols, predicate symbols, and constants. This way, we can rule out a harmful modal conservative extension of T as counting as a definitional extension, while allowing for harmless and trivial cases, such as the one above. And

the modification is continuous with the *spirit* of CDE, since it requires that new logical terminology be definable in just the same way as predicates and constants are. So let's drop the assumption that our logical constants must be held fixed, and add the following two conditions to definitional extension, to get definitional extension*:

$T+$ is a *definitional extension** of T iff:

- a) $T+$ is a definitional extension of T .
- b) For every operator O of $L+$, every formula $\phi+$ of $L+$, there is a formula ϕ of L such that:
 $T+ \models \forall x [O\phi+(x) \leftrightarrow \phi(x)]$.
- c) For every two-place connective $*$, every formula $\phi+$, $\psi+$ of $L+$, there's a formula ϕ of L such that:
 $T+ \models \forall x, y [(\phi+(x)*\psi+(y)) \leftrightarrow \phi(x, y)]$.
- d) For every quantifier Q of $L+$, every formula $\phi+$ of $L+$, there is a formula ϕ of L such that:
 $T+ \models \forall x [Qx [\phi+(x)] \leftrightarrow \phi(x)]$.⁵⁸

These conditions allow for the introduction of “dummy” operators: there can be a “modal” definitional extension of a non-modal theory, but only if we had a formulaic way to move

⁵⁸ A few notes here: first, I only list the criterion (c)) for two-place connectives, but it is easy to see how it would generalize to one, three, etc., place connectives. Second, note that these are meant to be axiom schemas. As such, it's possible that there are infinitely many axioms (e.g. could be one for each formula of $L+$) that satisfy the schemas, and one might worry that this is too weak. That is, there is no need for uniformity in the way we map sentences of $L+$ involving new logical terminology to sentences of L —there doesn't have to be a special kind of *form* that the corresponding L -sentence has, it needn't consistently contain the same logical constants (of L), and so on. This might seem too weak. But it is, in fact, intentional. Remember that we are only considering CDE* as a necessary condition for equivalence, and that I am going to argue that it rules out quantifier variance. As such, I *want* it to be fairly weak when it comes to logical cases. Besides, as we will see, the fact that we are dealing in uninterpreted theories will make a lot of these possible cases come out in the wash: we won't be able to generally get CDE*s where we have infinite axioms that satisfy the axiom schemas, since they'll be a mismatch between the kind of variation we get in interpretations of sentences that don't contain uniform logical constants and those that do.

Third, one also might want to consider here Barrett and Halvorson's (2015) discussion of *Morita Equivalence* as an alternative to extending definitional extension in the way I do here, particularly since it perhaps *would* allow for quantifier variance, though I think the details would have to be worked out there. I don't have the space to address taking Morita Equivalence to be a criterion for metaphysical equivalence here. But it is an interesting (and also similarly moderate) proposal.

between modalized sentences and truth-conditionally equivalent non-modalized sentences. They also allow for the introduction of connectives, so long as there is a formulaic way to get from sentences that involve the new connective to sentences that only involve the old one. These are exactly the right results. We can now allow for trivial cases of equivalence, such as the case involving '&' and 'v' I gave above, without losing any of the virtues of CDE outlined in sections 3 and 4.

6. CDE* Rules Out Quantifier Variance

One might worry that CDE* is too weak to give us much methodological guidance. But CDE* is in fact a very strong necessary condition, one which gives us some interesting results. One of these results is that CDE* rules out *quantifier variance*.

Quantifier variance is the claim that there are multiple distinct but *equally joint-carving* quantifiers. In the recent literature about the status of the composition dispute, many accept that quantifier variance is *necessary* for it to be the case that the nihilist and the universalist are not genuinely disagreeing. Hirsch (2005, 2009) takes quantifier variance to be true; Sider (2009, 2011) takes it to be false, and takes its falsity to establish that the dispute is meaningful. The idea is that the nihilist and the universalist must be using different quantifiers (or else we can't account for their prima facie disagreement about the status of certain sentences of pure logic, e.g. the claim that there are at least three things at a world with two atomic simples). But it can't just be that, for example, the nihilist is merely using a *restricted version* of the universalist's quantifier. If that were the case, then the nihilist, when faced with the universalist's sentence “there exists a table”, would have to admit that the universalist was quantifying over a domain that included tables. And then, if universalism and nihilism are equivalent—if they both are true

of the world, in virtue of using quantifiers with distinct meanings—the nihilist is forced to accept that there is domain which contains tables, and then is forced to accept that tables exist. So neither quantifier can be a mere restriction on another, and they both must be equally joint-carving.

But note that the quantifiers occupy the exact same proof-theoretic role for the universalist and the nihilist. Both the nihilist and the universalist will themselves *say* that they are using our regular old semantics and inference rules for their quantifiers.

And if they are right that they are each using our regular old semantics and inference rules for their quantifiers, then there is a very quick argument that there can be no CDE* of universalism and nihilism. Let's call universalism T_u and nihilism T_n . The quantifiers, exists_u in L_u and exists_n in L_n , have the same inference rules. When we move to L^+ , we cannot change those inference rules, for doing so would certainly amount to the introduction of a new constant and the elimination of an old one, violating the requirement that L^+ be a superset of L_u (and of L_n). I suppose one could resist this claim, but a quick reductio shows that resistance will lead to a violation of the conservative extension requirement. Suppose T^+ is a conservative extension of T_n . If we non-trivially change the inference rules for exists_n when we move from T_n to T^+ , then either a) we will be able to obtain as theorems of T^+ sentences of L_n that are not theorems of T_n (if our new rule allows us to derive *more* than we previously could), or b) we will not be able to obtain as theorems of T^+ some sentences of L_n that are theorems of T_n (if our new rule allows us to derive *less* than we previously could). Either a) or b) gets us that T^+ is indeed not a conservative extension of T_n .

Now suppose that T^+ is a CDE* of T_n and T_u . In light of the fact that they have the same inference rules, exists_u and exists_n will be interderivable (we just eliminate exists_n and then

introduce exists_u, or vice-versa). But now suppose that T_u says that there exists_u something composed of some particles arranged tablewise. It follows that T₊ says that there exists_u something composed of some particles arranged tablewise. But now in T₊ we can derive that there exists_n something composed of some particles arranged tablewise, which is false in T_n. So T₊ is not a definitional extension* of T_n, and hence is not a CDE* of universalism and nihilism.

This argument looks very much like (though not exactly like) one which Turner (2010) refers to as the *there can be only one* argument. Versions of it appear, for various reasons, first in Harris (1982) and also in Williamson (1988, 2006), McGee (2000, 2006). Turner's response, on behalf of the ontological pluralist, is to suggest that she adopt the free logician's inference rules for quantifiers. Turner's suggestion there is: (a) the inference rules for the quantifiers will not allow for interderivability of the quantifiers in the way I outlined above (since the introduction and elimination rules for each quantifier will require existential commitment *by the lights of that quantifier*), and (b) that this is a natural, not an ad-hoc, move to be made: the right way to understand pluralism is adopt this understanding of the quantifiers (to allow for multiple “ways of being”).

But this move comes with a great theoretical cost. It's true that if we move to free logic, we won't be able to interderive the universalist's and the nihilist's quantifiers, and positive model theory for free logic (model theory in which claims in the “outer domain”, which have no referent, can come out true) will allow the nihilist to speak truly about the universalist's commitment without committing herself to those entities. (So she can have a predicate, e.g., 'C', that she interprets as 'is a composite object'. Nothing in the inner domain of the nihilist's quantifier will fall under the predicate, but she has non-referential names, and can get claims like 'Ct' to come out true even when 't' doesn't refer.) This is desirable, since it allows the universalist

and nihilist to talk about each others' views without existentially committing to one another's ontologies. And we might think that we can now generate a CDE* that defines the universalist's quantifier as the disjunction of the nihilist's 'C' predicate and her own quantifier. But in fact we still can't do so.

Presumably, the universalist's quantifier would have to be defined as follows in T+:

$$T+ \models \forall x (E_u!x \leftrightarrow (E_n!x \vee C_n x)).$$

This claim is true on our *intended* interpretation of T+: something exists universally just in case either it exists nihilistically or is a composite object. But there are other interpretations on which it is false. For example, consider an interpretation on which we interpret 'C' as 'is a unicorn'. Such an interpretation is one in which *both* the nihilist and the universalist want it to come out false that whatever is in the extension of C is in the domain of their quantifiers. But then our proposed definition of the universalist's quantifier in T+ can't be right. It commits the universalist to quantifying over unicorns.

So we can't simply define the universalist's quantifier with the disjunction of the nihilist's quantifier and the 'C' predicate, for then the universalist will be committed to quantifying over unicorns. And there is no principled reason to ban interpretations like the unicorn one; the whole point of moving to free logic is to allow for the nihilist to talk about entities she takes to be fictional, and so it would be ad-hoc for us to restrict her to being able to talk about only the *universalist's* objects in such a way, or to prevent the universalist from being able to talk about fictional objects. I won't show this here, but I am fairly confident that this issue will extend to *any* non-logical vocabulary we try to supply the nihilist with to define up the universalist's quantifier, except perhaps in cases where we are dealing in implausibly simplified theories. Moving to free quantifiers can't help the quantifier variantist.

It follows that *if* quantifier variance is the only route to genuine deflationism about the dispute over composition, then CDE* entails that deflationism about composition is false. This is an interesting result. I motivated CDE* as a necessary condition for equivalence by claiming that there needs to be *some* linguistic structure (“perspective”) from which the two theories look identical and which preserves the structure of each theory, and by pointing to some sensible things that Chalmers says about verbal disputes. CDE* as necessary for equivalence does not obviously rule out quantifier variance. Importantly, it doesn't rule out the general possibility of there being multiple, equally joint-carving descriptions of the world, nor does it require that there is a single *most* joint-carving description of the world. This is because there might be two perfectly joint-carving descriptions of the world that have a CDE*. Treating CDE* as necessary for equivalence does not rule this out, because there is no requirement that our T+ be as joint-carving or more joint-carving than our T and T'.

For example, the theory that describes the world in terms of 'grue' and 'bleen' and the theory that describes the world in terms of 'blue' and 'green' theory may well have a CDE*, and hence satisfy my proposed necessary condition for equivalence. Are they equivalent? This is not settled by what I say here, since I have not proposed any sufficient conditions for equivalence. What is notable is that even anyone who thinks that talk of natural properties is nonsense, or that there is no sense to be made of saying that the green/blue theory of the world is “better” (in some metaphysical way) than the grue/bleen theory of the world is, can accept CDE* as a necessary condition for theoretical equivalence. It is in this sense a weak condition, and can be embraced by a broad spectrum of philosophers. The fact that it seems to rule out the thesis of quantifier variance, as construed by Hirsch and Sider, is hence of no small importance.

There are two things to note here. First, Sider (2009, 2011) argues against quantifier

variance by doing heavy-duty metaphysics. He both extends the notion of Lewisian naturalness, and argues that there *is* a single best (and most joint-carving) description of the world. But I have shown that there is another way to reject quantifier variance that rests on much less heavy-duty metaphysics, and which will be much more palatable to more moderate metaphysical realists.

Second, despite his self-proclaimed metaphysical realism, Hirsch has been accused, e.g. by Hawthorne (2009), of holding a position that can only be consistently motivated by verificationist principles. It has always struck me that Hawthorne is right about this. I hope, then, to be supporting Hawthorne's point. The proponent of quantifier variance will be forced to reject CDE* as necessary for equivalence. But even the most conservative metaphysical realist should, I think, either endorse CDE* or some *alternative* necessary condition for equivalence that respects our “perspective” talk. I hope to then at least have shifted the burden here: The quantifier variantist must tell us what, exactly, is wrong with *this* criterion of equivalence, which seems entirely metaphysically harmless and will be endorsed even by some philosophers of science who are entirely skeptical of metaphysics in the first place. And, if she wants to maintain that she is a realist, she must elaborate some *replacement* criterion for equivalence.

The view I have outlined here gives us a way to block quantifier variance without almost any of Sider's commitments. But it's not just Siderian commitments we needn't take on. We needn't be committed to any view on which there is a privileged class of natural properties to hold that CDE* is necessary for theoretical equivalence. (It is once we consider whether CDE* is sufficient for equivalence that we will be faced with these issues—but that is a question for another time.) It is interesting—and, I think, virtuous—that treating CDE* as necessary for equivalence is that it blocks quantifier variance without much heavy-duty metaphysics at all.

5. References

- Armstrong, D. (1978). *Universals and Scientific Realism, v. 1: Nominalism and Realism*. Cambridge University Press.
- Armstrong, D. (1986). In Defence of Structural Universals. *Australasian Journal of Philosophy* 64, 85-88.
- Armstrong, D. (1989). *Universals: An Opinionated Introduction*. Westview Press.
- Armstrong, D. (2004). *Truth and Truthmakers*. Cambridge University Press.
- Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics*. Oxford University Press.
- Barrett, T.W.. (2015). On the Structure of Classical Mechanics. *British Journal for the Philosophy of Science* 66, 801-828.
- Barrett, T.W. and Halvorson, H. (forthcoming). Glymour and Quine on Theoretical Equivalence, *Journal of Philosophical Logic*.
- Barrett, T.W. and Halvorson, H.. (ms). Morita Equivalence.
- Beall, J.C., and Restall, G.. (2013). Logical Consequence. *Stanford Encyclopedia of Philosophy* entry.
- Bennett, K. (2009) 'Composition, Colocation, and Metaontology', in Chalmers, Manley, Wasserman (eds.), *Metametaphysics*, Oxford University Press.
- Benacerraf, P.. (1964). What numbers could not be. *Philosophical Review* 74 (1), 47-73.
- Bhogal, H., and Perry, Z.P.. (2015). What the Humean Should Say about Entanglement. *Nous* 49(3).
- Boghossian, P. (1997). Analyticity. In Hale, B., Wright, C. (eds.) *A Companion to the Philosophy of Language*. Blackwell Press. 331–368.
- Chalmers, D.. (2011). Verbal Disputes. *Philosophical Review* 120(4), 515-566.
- Cowling, S.. (2013). Ideological Parsimony. *Synthese* 190 (17), 3889-3908.
- Dasgupta, S.. (2009). Individuals: An Essay in Revisionary Metaphysics. *Philosophical Studies* 145 (1), 35-67.
- Dasgupta, S. and Turner, J., Postscript. (2016). In Barnes, E. (ed.), *Current Controversies in Metaphysics*, Routledge.

- Dasgupta, S.. (2015). Substantivalism vs. Relationalism about Space in Classical Physics. *Philosophy Compass* 10(9), 601-624.
- de Bouvere, K.L.. (1965). .Synonymous Theories. in Atkinson, ed., *Symposium on the Theory of Models*. North Holland.
- Donaldson, T. (2015). Reading the Book of the World. *Philosophical Studies* 172(4), 1051-1077.
- Dorr, C. (2004). Non-symmetric Relations. *Oxford Studies in Metaphysics* 1, 155-192.
- Dresner, Eli. (2010). Logical Consequence and First-order Soundness and Completeness: A Bottom Up Approach. *Notre Dame Journal of Formal Logic* 52(1), 75-93.
- Dummett, M. (1978). Is logic empirical? In *Truth and Other Enigmas*. Harvard University Press, 269–289.
- Dutilh Novaes, C. and Andrade-Lotero, E. (2012). Validity, the Squeezing Argument, and Alternative Semantic Systems: the Case of Aristotelian Syllogistic. *Journal of Philosophical Logic* 41(2), 387-418.
- Eddon, M. (2013). Fundamental Properties of Fundamental Properties. *Oxford Studies in Metaphysics* 8, 78-104.
- Etchemendy, J. (1990). *The Concept of Logical Consequence*. Harvard University Press.
- Field, H. (1980). *Science without Numbers*. Princeton University Press.
- Field, H. (2009). What is the Normative Role of Logic?. *Aristotelian Society Supplementary Volume* 83(1), 251-268.
- Fine, K. (1994a). Essence and Modality. *Philosophical Perspectives* 8, 1-16.
- Fine, K. (1994b). Senses of Essence. In Sinnott-Armstrong, Raffman & Asher (eds.), *Modality, Morality, and Belief. Essays in Honor of Ruth Barcan Marcus*, Cambridge University Press.
- Fine, K. (2000). Neutral Relations. *Philosophical Review* 109(1), 1-33.
- Glymour, C.. (1970). Theoretical Realism and Theoretical Equivalence. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 275-288.
- Glymour, C.. (1977). The Epistemology of Geometry. *Nous* 11(3), 227-251.
- Glymour, C.. (2013). Theoretical Equivalence and the Semantic View of Theories. *Philosophy of Science* 80(2), 286-97.

- Hale, B.. (1987). *Abstract Objects*. Basil Blackwell.
- Halvorson, H.. (2012). What Scientific Theories Could not Be. *Philosophy of Science* 79(2), 183-206
- Halvorson, H.. (2013). The Semantic View, if Plausible, is Syntactic. *Philosophy of Science* 80(3), 475-478.
- Harman, G.. (1986). *Change in View: Principles of Reasoning*. MIT Press.
- Harris, J.H.. (1982). What's so Logical About the Logical Axioms. *Studia Logica* 41, 159-171.
- Hawthorne, J.. (2009). Superficialism in Ontology. In Chalmers, Manley, and Wasserman (eds.). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.
- Hicks, M. and van Elswyk, P.. (2015). Humean Laws and Circular Explanation. *Philosophical Studies* 172(2), 433-443.
- Hilbert, D. (1980). Letter to Frege. In Frege, G.. *Philosophical and Mathematical Correspondence*. University of Chicago Press.
- Hirsch, E. (2002). Quantifier Variance and Realism. *Philosophical Issues* 12, 52-73.
- Hirsch, Eli. (2005). Physical-object ontology, verbal disputes, and common sense. *Philosophy and Phenomenological Research* 70, 67-97.
- Hirsch, E. (2009). Ontology and Alternative Languages. In Chalmers, Manley, and Wasserman (eds.). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University \ Press.
- Hofweber, T. (2009). Ambitious, Yet Modest, Metaphysics. In Chalmers, Manley, and Wasserman (eds.). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.
- Jubien, M. (2001). Propositions and the Objects of Thought. *Philosophical Studies* 103, 47-62.
- Kant, I. (1999). (Guyer, P., and Wood, A., translators/eds.), *The Critique of Pure Reason*. Cambridge University Press.
- Krantz, Luce, Suppes and Tversk. (2006). *Foundations of Measurement Vol. I*. Dover.
- Kreisel, G.. (1972). Informal rigour and completeness proofs. In Lakatos, ed., *Problems in the Philosophy of Mathematics*. North-Holland. 138–171.
- Korman, D.Z.. (2010). Strange Kinds, Familiar Kinds, and the Charge of Arbitrariness. *Oxford*

Studies in Metaphysics, 119-144.

Koslow, A.. (1992). *A Structuralist Theory of Logic*. Cambridge University Press.

Lange, M.. (2013). Grounding, Scientific Explanation, and Humean Laws. *Philosophical Studies* 164(1), 255-261.

Lewis, D.. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61(4).

Lewis, D.. (1984). Putnam's Paradox. *Australasian Journal of Philosophy* 62(3), 221 – 236.

Lewis, D.. (1986). *On the Plurality of Worlds*, Oxford (Blackwell), 1986.

Loewer, B.. (2012). Two Accounts of Laws and Time. *Philosophical Studies* 160, 115-137.

Lutz, S.. (2015). What Was the Syntax-Semantics Debate in the Philosophy of Science About?. *Philosophy and Phenomenological Research* 91(2).

Maddy, P.. (2002). A Naturalistic Look at Logic. *Proceedings and Addresses of the American Philosophical Association* 76(2), 61-90.

Maudlin, T.. (1993). Buckets of Water and Waves of Space: Why Spacetime is Probably a Substance. *Philosophy of Science* 60(2), 183-203.

Maudlin, T.. (2012). *Philosophy of Physics: Space and Time*, Princeton University Press.

McGee, V.. (2000). Everything. In Sher and Tieszen, eds., *Between Logic and Intuition*. Cambridge University Press, 54-78.

McGee, V.. (2006). A Rule for Everything. In Rayo and Uzquiano, eds. *Absolute Generality*. Oxford University Press.

Merricks, T.. (2003). The End of Counterpart Theory. *Journal of Philosophy* 100, 521-49.

Miller, E.. (2015). Humean Scientific Explanation. *Philosophical Studies* 172, 1311-1332.

Miller, K.. (2005a). The Metaphysical Equivalence of Three and Four-Dimensionalism. *Erkenntnis* 62(1), 91-117.

Miller, K.. (2005b). What is Metaphysical Equivalence?. *Philosophical Papers* 34(1), 45-74.

Miller, K.. (ms) Metaphysical Equivalence and a Heterogeneous Account of Hyperintensionality.

North, J.. (2009). The 'Structure' of Physics: A Case Study. *Journal of Philosophy* 106, 57-88.

- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Peacocke, C.. (1988). The Limits of Intelligibility: A Post-Verificationist Proposal. *The Philosophical Review* 97(4), 463-496.
- Peacocke, C.. (1992). *A Study of Concepts*. MIT Press.
- Prawitz, D.. (1977). Meaning and proofs: on the conflict between classical and intuitionistic logic. *Theoria* 43, 2–40.
- Priest, G. (2014). Revising logic. In Rush, P., ed., *The Metaphysics of Logic*. Cambridge University Press, 211–223.
- Putnam, H. (1969). Is logic empirical? *Boston Studies in the Philosophy of Science* 5, 216–.
- Putnam, H. (1976). The logic of quantum mechanics. In *Mathematics, Matter and Method*. Cambridge University Press, 174–197.
- Quine, W.V.O.. (1948). On What There Is. *Review of Metaphysics* 2, 21–38.
- Quine, W.V.O.. (1951). Two Dogmas of Empiricism. *Philosophical Review* 60, 20-43.
- Rayo, A.. (2013). *The Construction of Logical Space*, Oxford University Press.
- Rosen, G. (ms).. Real Definition.
- Rosen, G. and Dorr, C.. (2002). Composition as a Fiction. In Gale, R. (ed.), *The Blackwell Companion to Metaphysics*. Blackwell, 151-174.
- Russell, G. K.. (2014). Metaphysical analyticity and the epistemology of logic. *Philosophical Studies* 171(1), 161–175.
- Russell, J.S.. (ms). Quality and Quantifiers.
- Schaffer, J.. (2010).. Monism: The Priority of the Whole. *Philosophical Review* 119 (1), 31-76.
- Sider, T.. (1996a). Naturalness and Arbitrariness. *Philosophical Studies* 81, 283-301.
- Sider, T. (1996b). Intrinsic Properties. *Philosophical Studies* 83(1), 1-27.
- Sider, T.. (2009). Ontological Realism. In Chalmers, Manley, and Wasserman (eds.). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.
- Sider, T.. (2011). *Writing the Book of the World*. Oxford University Press.
- Sklar, L.. (1974). *Space, Time, and Spacetime*. University of California Press.

- Suppes, P. and Zinnes, J.. (1963). Basic Measurement Theory. In Luce, R., Bush, R., and Galanter, eds. *Handbook of Mathematical Psychology*. Wiley.
- Turner, J.. (2010). Ontological Pluralism. *Journal of Philosophy* 107(1), 5-34.
- Turner, J.. (2012). Logic and Ontological Pluralism. *The Journal of Philosophical Logic* 41(2), 419-448.
- Turner, J.. (2016). *The Facts in Logical Space*, Oxford University Press.
- Van Fraassen, B.. (1980). *The Scientific Image*. Oxford University Press.
- Wang, J.. (ms). Fundamentality and Modal Freedom.
- Williamson, T.. (1985). Converse Relations. *The Philosophical Review* 94, 249–262.
- Williamson, T.. (1988). Equivocation and Existence. *Proceedings of the Aristotelian Society* 88, 109-127.
- Williamson, T.. (1994). *Vagueness*. Routledge.
- Williamson, T.. (2006). Absolute Identity and Absolute Generality. In Rayo and Uzquiano, eds., *Absolute Generality*, Oxford University Press.
- Williamson, T.. (2013a). *Modal Logic as Metaphysics*. Oxford University Press.
- Williamson, T.. (2013b). Logic, metalogic and neutrality. *Erkenntnis* 79(2), 211–231.
- Williamson, T.. (2015). Semantic paradoxes and abductive methodology. In Armour-Garb, B., ed., *The Relevance of the Liar*. Oxford University Press.
- Wittgenstein, L.. (1998). *Tractatus-Logico-Philosophicus*, Ogden (translator), 471st edition, Dover.
- Wright, C.. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen University Press.
- Wright, C.. (2001). On Being in a Quandary. *Mind* 110, 45-98.