AN EPISTEMIC ACCOUNT OF METAPHYSICAL EQUIVALENCE[1]

Michaela Markham McSweeney
Boston University

Introduction

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 other, they are rarely posed in the following way: what does it take for two theories to be equivalent? More often, the question of equivalence, while it underpins debates about verbal disputes and about the substantivity of certain metaphysical theses, is left lurking in the background. The aim of this paper is to tackle the equivalence question head-on.

Instead of giving an *analysis* of equivalence, though, I will argue for a necessary condition for justification in believing that two theories are equivalent. That is, this paper is about the *epistemology* of metaphysical equivalence. The proposal is as follows: in order to be justified in believing that two theories, T and T', are equivalent, there must be an occupiable perspective from which T and T' can be conceived of as a single unified theory, T+, which (in some to-be-determined sense) says nothing over and above either T or T', and which *says everything* that T does and T' does. Call this the **Unified Perspective Condition**. I will clarify and argue for the condition, as well as motivate the epistemic approach to metaphysical equivalence, in what is to come.

First, though, I need to say more about what metaphysical equivalence is and why we should care about it. We have some idea of what it is for two theories to be *empirically equivalent.* Some philosophers (verificationists, for example) will take metaphysical equivalence to just be empirical equivalence. This paper targets those who believe that there can be genuine distinctions between empirically equivalent possibilities (say, between a physicalistic world and its empirically equivalent dualistic counterpart). On such a view, it must take something more than mere empirical equivalence for two theories to count as metaphysically equivalent.

One might think that if two theories mean different things, they must not be metaphysically equivalent. But metaphysical equivalence is weaker than *meaning-equivalence.* There is some genuine distinction between the meaning of the sentence 'Annie is shorter than Kareem' and the meaning of the sentence 'Kareem is taller than

Annie' (whether it is a difference in propositions expressed is a more delicate question).[2] But it is plausible that these two sentences are equivalent with respect to all the worldly facts. They say the the same thing about the world, and so are metaphysically equivalent. So meaning inequivalence does not entail metaphysical inequivalence. Metaphysical equivalence hence requires *more* than empirical equivalence, but *less* than meaning equivalence.

What are some plausible candidates for being metaphysically equivalent theories? Perhaps: two theories that are exactly the same in content except that one uses '&' and '~' whereas the other uses 'v' and '~'; the two theories about Annie and Kareem's relative heights mentioned above; two theories that are exactly the same in content except that one uses the *gram* as a mass unit and the other uses the *shram* (where a *shram* = 1.4 grams). Thinking about what it would be to *deny* these equivalences helps us see what metaphysical equivalence is. For example, in the mass case, it would be to claim that the gram was more *metaphysically privileged* than the shram (or vice-versa); that it somehow 'carved nature at its joints' better than the shram did; that something about the world *decided between* the gram theory and the shram theory.

This is all very rough. Still, it seems clear why it matters. The question of when a given metaphysical dispute is *substantive* seems to depend on whether the competing theories in question are genuinely metaphysically inequivalent. Once I motivate the Unified Perspective Condition, I will show that it leads to some interesting results for certain metaphysical disputes. On one natural way of formalizing the condition, it rules out *quantifier variance,* the thesis that there are multiple, equally fundamental quantifiers. Some philosophers think that we need quantifier variance to show that the dispute between, e.g., the nihilist and universalist about composition is non-substantive. So if purported equivalences that employ quantifier variance can't meet the condition, then we can, perhaps, conclude that these disputes are indeed substantive.

While the quantifier variantist might reject this particular way of formalizing the condition, doing so will leave her with the burden of finding a different way to satisfy it, and so she faces a serious challenge. This is particularly notable because nothing about the Unified Perspective Condition presupposes that there is any sense to be made of metaphysical concepts like naturalness, fundamentality, or "levels" of reality. It seems to follow that metaphysical realists don't need to invoke these concepts, which many of us are worried by, in order to respond to the quantifier variantist.

The structure of the paper is as follows: In section one I argue for the Unified Perspective Condition. I also show that the condition should be attractive to a much wider range of philosophers—including neocarnapians—than more metaphysical accounts of equivalence are. In section two, I outline and briefly motivate one way of formalizing the

---

2   One accidental feature of this example is that these two sentences might mutually analytically entail one another, and it's important to note that it is not just cases like this that have a distinction in meaning without a distinction in the world. Consider 'Hesperus is shining' and 'Phosphorus is shining', for example.

Unified Perspective Condition. In section three, I argue that this formalization rules out *quantifier variance.* Finally, in section four, I make some remarks about how, and whether, the quantifier variantist could satisfy the Unified Perspective Condition some other way.

One somewhat unusual feature of this paper is that, while I mostly discuss the metaphysical equivalence *of metaphysical theories,* many of the tools I use come from the philosophy of science and might be taken to more properly apply to scientific theories. This is justified if we endorse even minimal metaphysical and scientific realism: equivalence is about whether there are any worldly differences between theories, and it should be uncontroversial that, while metaphysicians, scientists and philosophers of science might be talking about different aspects of the world, we're all talking about the same world. So we have no reason to think that the right account of equivalence could vary across these domains of inquiry.[3]

1. Motivating the UPC

Kristie Miller (2005a, 2005b) argues that two theories are equivalent just in case they share the same truthmakers. Theodore Sider (2011, 2016) tentatively argues for the claim that two theories are equivalent just in case they bottom out in the same fundamental concepts—they say the same thing about the world at the fundamental level. One similarity between the proposals is this: they both take seriously the idea that metaphysical equivalence is *equivalence by the lights of the world;* that two theories are inequivalent just in case there is some worldly distinction between them. Miller's and Sider's views are attractive. When are two theories equivalent? When, and only when, the world doesn't distinguish between them.

I want to think about constraints on and accounts of equivalence in terms of *perspectives.* What is a perspective? Let's say that a perspective is a way of conceiving of some situation or world or entity. A perspective is a way of conceiving of a *target.* (In many cases, the target will just be the world.) Sider's account require that, for two theories to be equivalent, there must be a perspective from which they are unified—from which they say a single, consistent collection of things about the world. And this perspective must be *metaphysically privileged*—it must be the perspective which views the world, or fundamental reality, as it really is. It must be completely unbiased, and it must not contain anything conventional, anything not metaphysically laden.

I am going to remain neutral on the question of whether this is really what it is for two theories to be equivalent, though given my realist sympathies, it is hard not to

---

3   For discussions of equivalence in the philosophy of science, in addition to Glymour (1970, 1977, 2013), Halvorson (2013, 2014), see, for example, Quine (1975) (though also Barrett and Halvorson 2016 for a discussion of how Quine's view relates to Glymour's). It is also worth examining more specific debates, e.g. North (2009), Halvorson (2011), and Barrett (2014) on whether Hamiltonian and Lagrangian mechanics are equivalent. (Though it is also important to note that it is unclear whether the parties to this debate are working with the same pre-theoretic notion of equivalence that I am.)

endorse this thought. What I doubt is whether *we could ever occupy* such a perspective—whether we could write down theories of the world that are completely transparent or unbiased. But even if one thinks we could do such a thing, it at least seems that we should grant how incredibly difficult it will be. So, if this is the right sort of account of equivalence, then it seems very hard, or impossible, to *learn* anything from it. The epistemic order of discovery goes from equivalence to 'same truthmakers' or 'same fundamental commitments', and not vice-versa.

In other words: these accounts might be right, but they aren't very informative. Whether we agree with the details, in broad brush strokes these theories just restate what we already took metaphysical equivalence to be. They don't help us make progress in deciding actual cases. And my goal is to think about equivalence in a way that will help us make progress in determining whether certain metaphysical disputes are substantive. In order to do so, we need to set aside the metaphysics of equivalence and focus on epistemology.

One obvious alternative to Sider and Miller-style accounts is that *nothing* is required for equivalence claims—that there needn't be any perspective at all from which the two theories are unified. I am not going to engage with this kind of account here, for it seems to me to be beyond the bounds of the very-roughly-realist starting point of this paper: at least as an epistemic claim, it seems to be unjustifiable unless we start with deeply anti-realist premises.

An apparent middle ground is to require there to be *some* perspective from which the two theories are unified, but not to require that perspective to be metaphysically privileged, or indeed privileged in any way at all. This is a good start. But for the realist, this middle ground is not nearly strong enough. It allows for very gerrymandered "unified" perspectives to get us equivalence claims. It makes no constraints on what our unified theory looks like, and we need these constraints. The Unified Perspective Condition has them built into it.

The ***Unified Perspective Condition*** (hereafter 'UPC'), recall, says that in order for us to be justified in claiming that two theories, T and T', are equivalent, there must be an occupiable *perspective* from which T and T' can be conceived of as a single unified theory, T+, which (in some to-be-determined sense) *says nothing over and above* either T or T', and which *says everything* that T and T' do.

This is a necessary condition on justification, and is not about when two theories are in fact equivalent. I suspect that providing an *informative* and perfectly general metaphysical account of metaphysical equivalence is beyond our limits. This is because there could be theories that are in fact equivalent that *we* don't have the resources (linguistic, formal, or otherwise) to unify into a single framework. But I don't know how we could formulate a criterion of equivalence that would both allow such pairs of theories to count as equivalent, and help us make judgments in cases that were not beyond our limits. And so I don't see how we could give a useful analysis of equivalence that included those pairs of theories—their equivalence is beyond our ken. Hence, again, my

focus on justification.[4]

Given a *particular* pair of theories, T and T', in order for us to be justified in thinking they are equivalent, there must be some perspective from which we can *view* them as equivalent. I am taking as a starting point the idea that two theories that seem to us to be *inequivalent* can be presumed so until reason to doubt this has been provided. If we can conceptualize a distinction between two theories, then the burden is on the person who claims those theories are equivalent to show us that that is the case. And a first step to doing so is to show us that we can *see* them as equivalent.

Why require that we be able to *see* the unification, or occupy the perspective from which T+ unifies T and T'--that is, why require that *we be able to see* that T+ says nothing over and above T or T', and says everything that they do? Because our question is about justification. Suppose I draw a circle on a piece of paper, point to it, and tell you that I have shown you the unified theory that demonstrates that T and T' are candidates for being equivalent. I have given you *no* reason for thinking that T and T' could actually be unified in the right way. You need to be able to see *how* my drawing of a circle unifies the two theories, and you need to be confident that it does so while retaining everything from each of them, and not adding anything to either of them. Even if I convinced you that my drawing of a circle was actually a perfect representation of fundamental reality, you *still* would not have a reason to think that my drawing unified T and T' in the right way. (This may seem obvious, but it is important for what is to come.)

What about the other two components of the UPC? I motivate them in sections 1.1 and 1.2. In section 1.3, I argue that, while I have thus far motivated the UPC by assuming that a fairly strong form of metaphysical realism is true, the condition is perfectly consistent with other metametaphysical views.

1.1 Why can't we remove structure?

Why must we require T+ to say *everything that T says and everything that T' says*? Consider the alternatives. If we allow ourselves to *delete* things (sentences, or bits of vocabulary, perhaps) from T and T' in constructing T+, we must be certain that in doing so, we aren't removing claims that make genuine metaphysical commitments. And if we allow ourselves to *replace* things, we must be certain that in doing so, we aren't changing

---

4 Why do I claim that the equivalence of theories that we don't have the resources to unify is beyond our ken? The following, at least in part: to believe that particular pairs of theories are equivalent without them meeting the condition is to believe in something that outstrips the bounds of conceptual possibility. Consider a (very imperfect!) analogy to metaphysical possibility: we might think that there are metaphysically possible entities (e.g. alien properties) that we can't conceptualize. And it might be perfectly reasonable to believe in such entities, generally. Likewise, we might have reason to think that there are pairs of equivalent theories that we can't *conceptualize as equivalent*. But it makes no sense to "name" a candidate entity that we can't conceptualize, and then *believe in that distinctive entity*. Likewise, it makes no sense to insist that two *particular* theories are equivalent if we can't conceptualize their target in such a way that they *both* successfully describe that world. And this is so *even if* there are in fact pairs of theories which are equivalent but which we can't conceptualize as equivalent.

the metaphysical commitments of our theories.

Consider a toy example: suppose T says '2 + 2 = 4' and 'some evolutionary theory is true', and T' says '2 + 2 = 4' and 'some creationist theory is true'. It would be a mistake to think that we can just delete the second sentence of each theory in order to generate a unified theory T+. And it would be a mistake to think that we can replace 'creationist' and 'evolutionary' with the same term in order to generate a unified theory T+.

In many cases, whether these bits of our theories make metaphysical commitments is exactly what is at stake in the first place; so we ought not help ourselves to the claim that they do not—for that stacks the deck in favor of equivalence claims. And I reiterate that one of my starting assumptions is that if two theories seem to be saying distinct things about the world, we need a defeater for that seeming in order to get an equivalence claim—our default setting should be inequivalence in most cases.

But what about cases where we seem to have solid reasons for thinking that particular bits of our theories are conventional, or non-metaphysically-committing? Why should we require, in *these* cases, that T+ contain everything from T and T'? In many such cases (e.g. the 'gram' and 'shram' theories mentioned above), it will be very easy to satisfy the UPC—that is, it will be very easy to show that there is some T+ that just has all the gram facts and all the shram facts in it, and gives us a way to unify them.

It's important to remember that T+ isn't supposed to be *metaphysically privileged* or *special* or even *appealing.* It isn't revealing our metaphysical commitments to us in any way. It doesn't matter that T+ will contain non-fundamental vocabulary and all sorts of conventional bits, and it doesn't matter if the theory that, in some sense, *really* unifies the 'gram' theory and the 'shram' theory is a *more fundamental* metaphysical theory that has built into it the correct metaphysics of quantity. We are not asking for the theory that really unifies our two theories. We are just asking for *a way to see* T and T' as unified, while putting some minimal constraints on what counts as unification.

We might worry that there will be some pairs of theories which are fundamentally equivalent (in Sider's sense) but which don't satisfy the UPC. However, if we already *know* that a given T and T' are fundamentally equivalent, *and* we accept Sider's proposal, clearly there is a way to make T and T' satisfy the UPC. We just put T and T' in 'fundamental form' before evaluating whether they satisfy the condition. Since their fundamental forms are identical, it is trivial to then construct a T+ (we just do nothing). And we are justified in restating them in their fundamental forms since we are accepting Sider's proposal.

If we *don't* know that a given theory-pair is fundamentally equivalent, we certainly can't follow some procedure that involves deleting structure from the theories in order to test for equivalence, since, by hypothesis, we don't know which commitments of the theories are fundamental or worldly, and which are not.

To sum up: if we don't know that T and T' are fundamentally equivalent, it must be that we don't know exactly what their fundamental commitments are. And if this is the case, we can't be justified in *removing* structure from T and T' in order to evaluate them

for equivalence, since we don't know which bits of them are metaphysically committing and which are not. Nor can we be justified in restating them in different terms, for doing so would require us to know what that restatement needed to preserve and what it did not. What we *are* justified in doing is checking to see whether we can extend them into a single theory that doesn't remove anything from either one'.

1.2 Why Must T+ Not 'Say Anything Over and Above' T or T'?
　　Why do we need the clause of the UPC that says that T+ must not *say anything over and above* either T or T'? Without this, it could be that T and T' are about entirely different topics, or that they have different targets. For example, T could be theory about how the various pigmentation patterns on butterflies' wings evolved, and T' could be theory about the psychology of infants. Now, suppose that both T and T' are true. They are clearly not metaphysically equivalent. We *don't* want them to meet the condition, for while it is only a necessary condition on equivalence, we want it to be informative and to not overgeneralize.
　　Without saying that T+ can't say *more* than either T or T', this pair of theories satisfies the UPC. Assuming that both theories are true, a complete theory of the world will say everything that each of them does, and it will certainly be a perspective from which they are unified—it gives us a way of seeing them as both true, as consistent with one another, and as part of the bigger picture. And we won't usually need a much grander theory to generate this problem. In some cases, we can simply *add* T and T' together to get a theory that is true, consistent, and which in some sense unifies T and T'. But a theory that says that infant psychology works a certain way *and* that the evolution of butterflies works a certain way is not any evidence at all that infant psychology working a certain way *is metaphysically equivalent to* the evolution of butterflies working a certain way.
　　So we need to restrict what T+ can add, not just to the *combination* of T and T', but to each theory itself: we need to show not just that T+ says nothing more than T and T' together, but that it says nothing more than *either one* does. The difficulty is in spelling out exactly what this means. I will describe one way of doing so in section 2. For now, I just hope to have motivated the need for a clause that does the work to block pairs of theories like the infant theory and the butterfly theory mentioned above from satisfying the UPC.

1.3 A Metametaphysical Step Back
Thus far I have motivated the UPC largely by taking on some very strong realist commitments. I have suggested that the condition is consistent with a picture on which there is a single, absolutely fundamental way the world is. But one needn't have these commitments in order to accept the condition.
　　The UPC doesn't assume that there is a single (perhaps absolutely fundamental), linguistic-framework-independent *way that reality is*. Instead, the UPC simply requires

that *we be able to make sense of* the unification of two theories. There will be almost no cases in which T+s will correspond to some fundamental way the world is, or some way we think the world is. Our T+s are going to be ugly, large, overcrowded, redundant things. They are just *tools* for us to make sense of two apparently conflicting, or at least distinct, theories at the same time.

Moreover, we don't even have to think that there *is* a single (perhaps fundamental) *way the world is* to accept that we need these tools. Those who reject that there is a single way the world is *still* should accept the need for T+s. For their role is *not* to show us anything about the world, but to show us that *we* can see two different theories as reconciled, unified, and not in tension with one another. And without this, we would immediately descend into the view that *nothing* is required for us to be justified in making equivalence claims.

Why does this matter, given that the motivation I've given thus far targets the hardcore realist? In section 3, I will show that one way of fleshing out the UPC rules out the thesis of *quantifier variance.* Quantifier variance says that there are multiple, equally fundamental existential quantifiers. Quantifier variance is sometimes taken to be necessary in order to establish that, e.g., universalists and nihilists about composition are having a merely verbal dispute. And on one way of thinking about quantifier variance, it is a view which *rejects* the idea that there is a single, absolutely fundamental, linguistic-framework-independent *way the world actually is.* If the UPC should only be endorsed by hardcore metaphysical realists who *do* think there is a single, absolutely fundamental way the world is, then it is uninteresting that the quantifier variantist can not meet the condition. But in fact, the condition can and should be endorsed by philosophers with a wide range of metametaphilosophical views.

Neocarnapians, for example, needn't reject the UPC. This is because it is consistent with questions of equivalence being *internal* questions—being questions that only make sense *within* a particular linguistic framework, rather than being external questions that attempt to transcend linguistic frameworks (e.g. "but are there *really* numbers?"). Indeed, the very spirit of the UPC can be thought of as, in some ways, neocarnapian: upon initial examination the question 'when are two theories equivalent' looks like a paradigmatic external question: it is asking us to "step outside" the bounds of the linguistic frameworks in which those theories are constructed in order to evaluate some deeper, worldly, metaphysical question. But the UPC doesn't see equivalence this way: it tells us that the only way to begin to evaluate two theories for equivalence is from *within* a linguistic framework (a perspective!). (Though remember that this is only, and only should be for the hardcore realist, a *necessary* condition for a justified belief in equivalence—perhaps some will want to treat it as also sufficient, but hardcore realists should not.)

The neocarnapian should clearly reject some of my earlier claims, and even, perhaps, the pre-theoretic gloss on equivalence I gave in the introduction. But she should not obviously reject *the actual condition*. After all, to satisfy the condition is just to be

able to construct a linguistic framework from which the question of equivalence can be answered: from which we can "see" theories in other linguistic frameworks, which are parts of our new one, as unified. And surely that is exactly what the neocarnapian needs in order to make sense of the question of equivalence in the first place. So insofar as she wants questions of equivalence to make sense at all, she should endorse the UPC.

22. Common Definitional Extension*

In the previous section, I motivated the UPC. I will now propose one way of understanding the condition that gives us a technical tool for evaluating whether pairs of theories meet it. The proposal is that in order for us to be justified in thinking that two first-order theories are equivalent, they must have a common definitional extension* ('CDE*'). The CDE* condition is adapted from Glymour (1970, 1977), who argues that two theories must have a common definitional extension ('CDE') in order to be equivalent (though perhaps not in virtue of satisfying the UPC). I want to make clear that I remain open to other ways of satisfying the UPC. But the proposal here is important for two reasons. First, it gives us a way to actually *test* whether theories meet the UPC. Second, as I'll argue in section 3, it delivers interesting results: it rules out quantifier variance, and hence (according to some) rules out equivalences between certain metaphysical theses (e.g. nihilism and universalism about composition). The question is then whether proponents of the equivalence of nihilism and universalism could satisfy the UPC some other way.

In order to explain what a CDE* is, I need to explain what a CDE is. And in order to do that, I need to explain 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. (So every theorem of T is a theorem of T+, and 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+ $\models \forall x \, (Rx \leftrightarrow \varphi(x))$
> b) for every function symbol f of L+ there's a formula φ of L such that:
> T+ $\models \forall x,y \, (f(x) = y \leftrightarrow \varphi(x,y))$
> c) for every constant symbol c of L+ there's a formula φ of L such that:
> T+ $\models \forall x \, ((x=c) \leftrightarrow \varphi(x))$
> (Note that in each of these cases, the L-formula need not "match" the L+

formula—so e.g. in a) φ(x) might be non-atomic.) formula—so e.g. in a) φ(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 vocabulary of L and L+ this will be trivial, but for non-shared vocabulary it is not.

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'. In a moment, I will suggest a minor problem with using CDE to satisfy the UPC, and then a tweak to fix the problem. But it is fairly clear that two theories' having a CDE roughly satisfies the condition, and spelling out why this is should help clarify CDE itself.

What does it take for a given T and T' to have a CDE, T+? First, T+ must *include everything in both T and T'* (nothing is *removed* from T or T' in constructing T+, because T+ conservatively extends each). This satisfies the clause of the UPC which says that our T+ must *say everything that T does and say everything that T' does.* Second, there is a clear sense in which T+ *says nothing more* than T does (and says nothing more than T' does): the definitional clauses (a)-(c) require that any new vocabulary in L+ (the language of T+) is *defined* in vocabulary that appears already in T.

Further, note that insofar as we can understand T and T', we will be able to understand T+, since we will have a translation procedure for translating anything "new" that T+ says back into the language of T or T'. This helps CDE satisfy the UPC as well: the UPC tells us not just that there must *be* some T+ that in some sense unifies T and T', but that we must be able to occupy the perspective from which we can *see* that T+ unifies T and T'. Again, I won't argue here that CDE is the *only* way to satisfy the UPC. But it is fairly clear that it captures the spirit of the condition, in a way that allows us to actually test pairs of theories. The UPC is, in a sense, the *motivation* for treating having a CDE as necessary for equivalence.

However, there is a problem with the claim that having a CDE satisfies the UPC: 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 components of our vocabulary are held fixed between the two theories: that every logical constant in L is in L', and vice-versa. We want to be able to evaluate theories that contain different logical constants for equivalence. For example, we want to be able to evaluate whether the '&' and '~' theory is equivalent to the 'v' and '~ theory, even if the two theories seem to us to be mere notational variants. And we also want to be able to evaluate two theories that seem to assume different background logics entirely. There are some very tricky issues here. I discuss two of them (why do I claim that the definition of CDE assumes that logical vocabulary is held fixed? What justifies the classical assumptions here?) in the appendix.

The way that CDE is set up assumes that logic will be held fixed across theories.

How do we fix this so that we can evaluate theories that have distinct logical vocabularies? The most non-ad-hoc way to do so is to drop the implicit assumption that the logic is held fixed, and add *explicit* conditions that require new logical constants to be defined in similar ways to function symbols, predicate symbols, and constants. By dropping the implicit assumption, we can allow for harmless and trivial cases, such as the one above, to have a CDE*. But unless we then add explicit conditions that new logical constants in T+ must be definable in the languages of L, L', many pairs of theories that are clearly not equivalent will end up meeting the condition.

The modification is consistent with the *spirit* of CDE, and with the UPC, since it requires that new logical terminology be definable in just the same way as predicates and constants are. Once we drop the assumption that our logical constants must be held fixed, we can define definitional extension*:

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

a) T+ is a definitional extension of T (*with the implicit assumption that T and T' share logical vocabulary dropped*).
b) For every operator *O* of L+, every formula φ+ of L+,there is a formula φ of L such that:
T+ |= ∀x [*O*φ+(x) ↔ φ(x)].
c) For every quantifier *Q* of L+, every formula φ+ of L+,there is a formula φ of L such that:
T+ |= ∀x [*Q*φ+(x) ↔ φ(x)].
d) For every two-place connective *, every formula φ+, ψ+ of L+, there's a formula φ of L+ such that:
T+ |= ∀x,y [(φ+(x)*ψ+(y)) ↔ φ(x,y)].[5]

T and T' have a ***common definitional extension**** **(CDE*)**, T+, if T+ is a definitional extension* of T, and T+ is a definitional extension* of T'. These conditions allow for the introduction of "dummy" operators: for example, there can be a modal definitional extension of a non-modal theory, but only if we have a formulaic way to move 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. CDE* allows for trivial cases of equivalence, such as the '&' and '~' theory and the 'v' and '~' theory, and we retain all the motivation for it being a good formalization of the UPC.

So it seems clear that one plausible way to flesh out the UPC is to claim that two

---

5   This list is incomplete, but it is easy enough to see how it would extend to three-place connectives, and so on.

theories must have a CDE* in order for us to be justified in thinking they are equivalent. Is CDE* the best way of doing so? I am not at all sure, but I think it is a good contender. Here is one objection: I've assumed, in this section, that the occupiable perspective from which we must be able to *see* T and T' as unified must be a *linguistic* one, and, moreover, that we can give an entirely *syntactic* specification of what counts as a T+. In section 2.1, I will respond to this objection.

2.1 Why Go Linguistic? Why Go Syntactic?
Why should we think that the perspective from which two theories look identical needs to be a linguistic perspective? Perhaps what matters is that there is a way to "see" the models of two theories as identical. (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, and perhaps we can do something resembling that.) Someone who holds the semantic view of theories (roughly, that theories are individuated by their models) 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 (very roughly, that theories are sets of sentences) is right, or, more accurately, that insofar as we can make sense of semantic views, they end up collapsing into syntactic views.[6] But I will not argue for that here. Instead, I will focus solely on the epistemic issues with the semantic approach, for obvious reasons.[7]

---

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

7   For those interested in potential semantic accounts of equivalence, see Halvorson 2012, 2013, Glymour 2013. To briefly summarize: One thing we 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 "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 to think of what isomorphisms *are.* But, following Halvorson (2012), 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. One might react to this by pointing out the results

12

Any purely semantic view of equivalence will be methodologically unhelpful, even if we grant that it is *true*. What we get, from scientists and philosophers, is usually written words and symbols on a page. The challenge is to say 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, 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... The space of models of, say, Newtonian theory cannot be 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 *epistemic* problem that this paper is about, because our theories would still be expressed in 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

When it comes to the epistemic question about equivalence, there is a strong similarity between the problem with semantic views of theories and the problem with treating equivalence as bottoming out in fundamentality or truthmakers. 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. The semantic view, like the fundamental equivalence view, 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 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

---

in model theory, which I will outline briefly in section 3, 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 Bouvère'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 Bouvère results if we already accept that theories are linguistic entities.

sentences. So the move to a syntactic characterization of the UPC is justified.

3. CDE* Rules Out Quantifier Variance

I have argued that, in order for us to be justified in believing that two theories, T and T', are equivalent, they must satisfy the UPC. And I have shown that having a CDE* is at least one plausible way that T and T' can satisfy the UPC. Given that the discussion is only about what is *necessary,* and not what is *sufficient,* for justification, none of this will be interesting unless CDE* is strong enough to give us methodological guidance. Here, I will show that it is strong: it 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 facia 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).

So if all this is right, and if I am right that CDE* rules out quantifier variance, then we can't be justified in believing that views like universalism and nihilism about composition are equivalent. And that is certainly a strong result!

For quantifier variance to be true, 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 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 the other, and both must be equally joint-carving.[8]

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 doing so, 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).[9]

---

8   Note that one needn't buy into "joint-carvingness" or "levels" of reality or anything like that to accept this.

9   I suppose one could resist this claim, but a quick reductio shows that resistance will lead to a violation

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$_u$ sentences and exists$_n$ sentences will be interderivable in L+ (we just eliminate exists$_n$ and then introduce exists$_u$, or vice-versa). If 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) *collapse arguments,* or arguments which Turner (2010) refers to as the *there can be only one* argument, which appear in Harris (1982) and also in Williamson (1988, 2006), McGee (2000, 2006), Dorr (2014). I can't consider all the responses to these arguments. But I do want to consider one of them: Turner's response to this argument applied to a related debate—the debate over ontological pluralism--is to suggest that the ontological pluralist adopt the free logician's inference rules for quantifiers. He argues that (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 we can't make this move in the context of endorsing CDE* without incurring a serious 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 committing to each other'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.

There are a number of ways we might attempt to define the universalist's

---

of the conservative extension requirement. Suppose T+ is a conservative extension of Tn. If we non-trivially change the inference rules for existsn when we move from Tn to T+, then either a) we will be able to obtain as theorems of T+ sentences of Ln that are not theorems of Tn (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 Ln that are theorems of Tn (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 Tn . So we must think of the inference rules as fixed.

quantifier in T+. Suppose we try the following definition:

$$\text{T+} \models \forall x\, (E_u!x \; ? \; (E_n!x \lor C_n x)).$$

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 define the universalist's quantifier with the disjunction of the nihilist's quantifier and the 'C' predicate, for that would commit the universalist to quantifying over unicorns. And there is no principled reason to ban interpretations like the unicorn one; the 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, and not other objects she takes to be fictional. It seems likely that this problem will extend to *any* non-logical vocabulary we try to supply the nihilist with to define up the universalist's quantifier, since it stems from the distinction between logical and non-logical vocabulary. Moving to free quantifiers can't help the quantifier variantist.

Of course, there are other routes the quantifier variantist might take. Perhaps she should insist on a move to plural quantification. My suspicion is that this will not resolve the issue, but I won't argue for that here. In general, what I want to emphasize is that treating CDE* as a necessary condition for a justified equivalence belief creates a context in which collapse arguments are much *harder* to resist than normal. Moves that might be more plausible in general, such as the move to free quantifiers, don't work here. And that is because if we haven't adopted the condition, we can resist the requirement for the kind of definitional equivalence it imposes. Since the condition forces us to provide definitions, in the form of theorems of T+, for the logical vocabulary of L and L', it will force us, much more quickly, into accepting these kinds of arguments.[10]

If quantifier variance is the only route to genuine deflationism about the dispute over composition, then CDE* entails that we shouldn't believe deflationism, or, at the least, that it is very hard to figure out how to justify such a belief. This is an interesting result. I motivated CDE* as necessary for a justified equivalence equivalence claim by first arguing for the UPC, and then showing that two theories having a CDE* is a good way to satisfy the condition. But, as I emphasized in section 1.3, the UPC does not

---

10 E.g. I suspect that some of the arguments in Dorr (2014) will not work in this context, but I do not want to discuss more responses to collapse arguments here, particularly since I think the right move for the quantifier variantist here is to reject CDE* to begin with. I just wanted to show why one such argument doesn't work in the context of CDE*

obviously rule out metaphysical deflationism at all. It doesn't rule out the general possibility of there being multiple, equally joint-carving descriptions of the world, nor even multiple, equally fundamental *ways the world is*. Does CDE* rule this out? No, since there is no requirement that our T+ be *as fundamental as* T and T'. Indeed, T+s will almost necessarily contain ugly redundancies. Instead, as with the UPC itself, what CDE* demands is merely a *way of seeing* T and T' as unified, a perspective from which they look unified, even if (as will almost always be the case), that perspective itself clearly is not 'worldly' or fundamental or even remotely theoretically attractive.

4. Quantifier Variance and the UPC
There are ways that the quantifier variantist could resist the argument in section 3. But I don't think that is her best option here (and the main proponent of quantifier variance, Hirsch, would clearly reject CDE* as a good necessary condition for justified equivalence beliefs anyway). She could also, of course, reject the UPC altogether. I hope to have shown that that is not an attractive option: the UPC is the minimum necessary condition we need for justification for an equivalence belief, at least if we are realists.

The only alternatives are either to treat empirical equivalence as enough, or to claim that we don't need *anything* to justify equivalence beliefs. But Hirsch, at least, would not like either of these views. He is a metaphysical realist (and indeed, quantifier variance is only particularly *interesting* as a realist view—it should come out trivial, or close to it, on many anti-realist views) and so he must think that *something* is required to justify an equivalence belief. And he is not (purportedly) a verificationist, so empirical equivalence can't be enough.[11]

The quantifier variantist should instead respond by claiming that CDE* is too strong, and that there are other, weaker ways of fulfilling the UPC. It is important to point out that CDE* as necessary for a justified equivalence belief is actually *not* very strong in one important sense: it *does not* rule out that theories like the blue/green theory of the world and the green/blue theory of the world are equivalent. At least in some form, these theories meet the necessary condition: assuming we have temporal vocabulary in each, it is easy to see that we can construct a third theory that simply contains all of 'grue', 'bleen', 'green', and 'blue', and has theorems that state definitions of terms from each theory in terms of the other.

So even the philosopher 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*. And more generally, CDE* doesn't have anything built into it that requires us to think in terms of fundamentality, more or less metaphysically privileged descriptions of the world, "levels" of reality, or there even being *a single way the world is*. It is in this sense very weak. The fact that it rules out quantifier variance, as construed

---

11 Though see Hawthorne (2009) for an argument that only verificationism can motivate Hirsch's views.

by Hirsch and Sider, is hence of no small importance, *even if* the quantifier variantist might want to reject it for other reasons. (And note that it is also of importance for the precursor to the Hirsch-Sider debate, Lewis' (1984) invocation of naturalness in response to Putnam's (1985) model-theoretic argument.)

Like the UPC itself, CDE* also doesn't *rule out* that fundamentality, "levels" of reality, metaphysically privileged descriptions of the world, and so on might matter very much to equivalence. Since both are meant to be merely *necessary,* but not *sufficient,* for justified equivalence beliefs, the proponents of metaphysical pictures in which these notions play a prominent role can simply build them into *further* requirements for justified equivalence beliefs. So both the UPC and CDE* are compatible with, but neutral with respect to, the importance of these "heavy-duty" metaphysical notions. CDE* is incompatible with quantifier variance. Is the UPC itself incompatible with quantifier variance?

It is unclear. There are two kinds of strategies the quantifier variantist might use to meet the UPC. First, she could reject altogether the idea that we need a formal fleshing out of the UPC, one which tells us how to test for whether we can construct a theory, T+, that will satisfy the condition. The problem is that it is unclear what alternatives could satisfy the UPC. The discussion in section 2.1 shows that a purely semantic alternative would be problematic, for it would not give the UPC the epistemic force it needs. And while it is certainly an option in logical space to claim that we could satisfy the UPC entirely non-linguistically, it is very hard, as I emphasized earlier, to see how that would go.

I suspect, though, that neither of these is what the quantifier variantist wants anyway. One remaining option is to claim that we can *directly observe* that the world satisfies both theories, and that that is enough to show that the theories satisfy the UPC. But this isn't right. For what we need to know is whether there is a single perspective *we* can occupy in which we can *see* the world as satisfying both theories simultaneously (so to speak), that is, in a unified way, not having to step out of the bounds of one theory in order to evaluate, for the other, whether the world satisfies it. We need to show that we can observe the world from a single linguistic and conceptual framework, and see, from *within* that framework, that the world satisfies both theories. Claiming that it is simply obvious that the world satisfies both theories, but refusing to give an account of *how,* is dangerously close to claiming that we don't need anything at all to justify equivalence claims between empirically adequate theories. (And we wouldn't want the world to be able to count as our T+ anyway—for, at best, that would be to make the UPC amount to nothing more than empirical equivalence.)

It is plausible, then, that the only way to genuinely satisfy the UPC is by showing that there is a *linguistic framework* from which we can do so. And it seems to follow that the quantifier variantist must embrace the general idea that we need *something like* CDE* in order to flesh out the condition, but reject the specifics. That burdens her with providing such an alternative. One strategy here—the most promising one, I think--would

be to insist on using a different logic, and a different conception of model theory, in the very construction of a condition like CDE*.

Why would the quantifier variantist need to do this? Focus on CDE for a moment: 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 (or, to be less question-begging, an isomorphic set of objects) and says a corresponding true thing about it (and vice-versa). I'll explain this in a moment, but for those tired of technical details, the point is that if two theories have a CDE (and, perhaps, a CDE*, though I haven't proven this), then they have isomorphic classes of models, where *pairs* of models of T, T' have the same number of elements in their domains. And this is exactly what the quantifier variantist will want to reject, since she must either allow for mappings between models of theories with *different* numbers of models in their domains, or reject the mapping-between-models requirement altogether.

De Bouvère (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+. What does this mean? 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 <M, M+> will have the same number of elements in their domains***.[12]

In other words: the green/blue theory and the grue/bleen theory have a CDE in part precisely because we hold fixed all the facts about how many individuals there are in stating each of these theories. We can vary interdefinable predicates all we want. But we can't vary the number of elements in our domain. Since this is precisely what the quantifier variantist wants, she must reject CDE (and CDE*). And it seems that she must reject *any* candidate condition that has similar model-theoretic commitments.

My goal here is not to show the quantifier variantist how to meet the UPC. The burden is on her to show how she can do so, and if she must do so technically, then she needs to provide a way to resist the "pairs of models have the same number of elements" feature of CDE. To figure out how to consistently do so, *in a way that we can make sense of* (for remember our 'occupiable perspective' requirement), while not trivializing the condition (that is, while not saying "we can just map models with any number of elements to one another without any constraint") will be very difficult. Perhaps moving to

---

12 These results are shown in de Bouvère (1965).

plural quantifiers in the formalization itself will help (I am suspicious). Perhaps a radically revised view of model theory will help. Perhaps category theory will help (though this would require, I think, a complete restatement of quantifier variance itself). Perhaps rejecting that isomorphisms are even relevant to equivalence will help. The point is that all these options are hard, and it is not clear how they would go, and whether, in the end, they would respect the UPC.

It is my sense that to some metaphysicians, quantifier variance—or, more accurately, the metaphysical picture associated with quantifier variance when applied to particular disputes, like composition--feels both deeply challenging, and simultaneously, in some sense, unintelligible. We might take this to show that the foray into quantifier variance, when it comes to questions like whether there are tables, is deeply mistaken. But setting that possibility aside, my suspicion is that the explanation for why this thesis feels unintelligible (even when it can be clearly stated) is that quantifier variance violates the UPC.

We might think that realists, but not anti-realists, must tell us *what the world is really like,* since, after all, only realists think there is a way that the world really is. The quantifier variantist is, I take it, trying to occupy realism without having to tell us what the world is really like. But even a minimal kind of realism requires that our equivalence claims meet the UPC. Meeting the UPC doesn't require anything close to telling us what the world is really like. All it requires is *giving us a way to conceive* of the equivalence of two theories, over and above just foot-stomping, which should be out of bounds for even minimal realists.

Perhaps I haven't got the condition quite right, or perhaps quantifier variance doesn't violate it. I want to leave both those possibilities open. But in order to be justified in making equivalence claims about theories—and, indeed, perhaps in order to even be able to make *intelligible* equivalence claims about theories—we need the UPC, or something close to it. And so the quantifier variantist, insofar as she wants to claim that she is a realist, can't simply ignore the UPC altogether. More generally, none of us can.

References

Barrett, Thomas. 2015. "On the Structure of Classical Mechanics." *The British Journal for the Philosophy of Science* 66:801-828.

Barrett, Thomas and Halvorson, Hans. Forthcoming. 'Glymour and Quine on Theoretical Equivalence', *Journal of Philosophical Logic.*

De Bouvère, K.L. 1965. "Synonymous Theories." In *Symposium on the Theory of Models,* edited by Atkinson. North Holland Publishing Company.

Dorr, Cian. 2014. "Quantifier Variance and the Collapse Theorems." *The Monist* 97:503-570.

Glymour, Clarke. 1970. "Theoretical Realism and Theoretical Equivalence." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 197):275-288.

Glymour, Clarke. 1977. "The Epistemology of Geometry." *Nous* 11(3):227-251.

Glymour, Clarke. 2013. "Theoretical Equivalence and the Semantic View of Theories." *Philosophy of Science* 80(2): 286-97.

Halvorson, Hans. 2012. "What Scientific Theories Could not Be." *Philosophy of Science* 79(2):183-206.

Halvorson, Hans. 2013. "The Semantic View, if Plausible, is Syntactic." *Philosophy of Science* 80(3):475-478.

Harris, J.H. 1982. "What's so Logical About the Logical Axioms?" *Studia Logica* 41:159-171.

Hawthorne, John. 2009. 'Superficialism in Ontology.' In *Metametaphysics: New Essays on the Foundations of Ontology,* edited by David Chalmers, David Manley, and Ryan Wasserman. Oxford: Oxford University Press.

Hirsch, Eli. 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, Eli. 2009. "Ontology and Alternative Languages." In *Metametaphysics: New Essays on the Foundations of Ontology,* edited by David Chalmers, David Manley, and Ryan Wasserman. Oxford: Oxford University Press.

Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61(4):343-377.

Lewis, David. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62(3):221-236.

Lutz, Sebastian. 2015. "What Was the Syntax-Semantics Debate in the Philosophy of Science About?" *Philosophy and Phenomenological Research* 91(2). Doi:10.1111/12221.

McGee, Vann. 2000. "Everything." In *Between Logic and Intuition,* edited by Sher and Tieszen. Cambridge: Cambridge University Press. 54-78.

McGee, Vann. 2006. "A Rule for Everything." In *Absolute Generality,* edited by Agustin Rayo and Gabriel Uzquiano. Oxford: Oxford University Press.

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

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

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

Putnam, Hilary. 1985. *Realism and Reason: Philosophical Papers, Volume 3.* Cambridge: Cambridge University Press.

Rosen, Gideon and Dorr, Cian. 2002. "Composition as Fiction." In *The Blackwell Guide*

to Metaphysics, edited by R.M. Gale. Oxford: Blackwell.

Sider, Theodore. 2009. "Ontological Realism." In *Metametaphysics: New Essays on the Foundations of Ontology,* edited by David Chalmers, David Manley, and Ryan Wasserman. Oxford: Oxford University Press.

Sider, Theodore. 2011. *Writing the Book of the World.* Oxford: Oxford University Press.

Sider, Theodore. 2016. "Theoretical Equivalence." Presented as Lecture 6 of the 2016 Locke Lectures, Oxford University.

Turner, Jason. 2010. "Ontological Pluralism." *Journal of Philosophy* 107(1):5-34.

Turner, Jason. 2012. "Logic and Ontological Pluralism." *Journal of Philosophical Logic* 41(2):419-448.

Van Fraassen, Bas. 1980. *The Scientific Image.* Oxford: Oxford University Press.

Williamson, Timothy. 1988. "Equivocation and Existence." *Proceedings of the Aristotelian Society* 88:109-127.

Williamson, Timothy. 2006. "Absolute Identity and Absolute Generality." In *Absolute Generality,* edited by Agustin Rayo and Gabriel Uzquiano. Oxford: Oxford University Press.

Appendix

In section 2, I suggested that we needed to move from CDE to CDE*. I want to make two remarks about this. First, I need to explain the claim that CDE contains a background assumption that logical constants have to be held fixed across theories that have CDEs. 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.

Does the presupposition that we have to hold our logic fixed when evaluating theories for a CDE *also* rule out some of the most obvious cases of equivalence? For example: suppose that the only logical constants in L are 'v', '∼', and the existential quantifier. And suppose that the only logical constants in L' are '&', '∼', 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. Does the presupposition stop us from simply expanding L to include '&' (or, equivalently, expanding L' to include 'v') 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.

I think it is, at the least, very unclear whether these theories will have a CDE. In order for them to have one, the background assumption has to be something like 'hold the logical constants fixed, but allow for differences in logical constants that are *logically equivalent* to what we already have'. This may be what is intended. But it still doesn't allow us to compare theories that are stated using different logics, and not just trivially distinct logical vocabularies.

However, this discussion raises a potentially serious problem with CDE*: it requires us to use classical logic to define equivalences between potentially non-classical theories, when the very point is, in part, to stay neutral about whether theories with different logical constants can be equivalent. This brings me to my second remark. In one sense there is no way around this problem (and it is a familiar one from the philosophy of logic): we have to choose some way to state the definitional requirements in CDE*, so we will be choosing sides in the logic of our metalanguage at least. But there are two mitigating factors in this case: first, I am *not* proposing that CDE* is the only way to flesh out the UPC, and am perfectly happy to grant that, so long as a logic is intelligible, we

can construct countless alternatives to CDE* by making different assumptions about logic. What matters, more, is that we be able to identify *some* way in which a pair of theories can meet the condition, and so it *is* true that in identifying that way we will have to pick sides, even if we are pluralists about the options for doing so.

       Second, it is important—because of the "occupiable perspective" component of the UPC—that we ensure that, however we do construct something like CDE*, it involves a logic—a single logic--that we can make sense of. So there are in fact boundaries to this pluralism. One way of seeing the claim that it is hard to see how quantifier variance could meet the UPC is this: it is hard to see how quantifier variance could meet *any* more formal statement of the condition, given the boundaries of this pluralism.