

Keeping up appearances: a reducer's guide¹

David Manley

University of Michigan, Ann Arbor

Metaphysicians with reductive theories of reality like to say how those theories account for ordinary usage and belief. A typical strategy is to offer theoretical sentences, often called ‘paraphrases’, to serve in place of various sentences that occur in ordinary talk. But how should we measure success in this endeavor? Those of us who undertake it usually have a vague set of theoretical desiderata in mind, but we rarely discuss them in detail. My purpose in this paper is to say exactly what they are, and why.

Among the questions I want to address: what counts as an adequate theoretical replacement for an ordinary sentence? On whom does the burden of proof lie when it is unclear whether a theory of reality can provide a true replacement for some ordinary truism? Which ordinary truisms are most important to ‘save’ in this way, and why? What role does the theoretical virtue of simplicity play in this strategy: is the goal to offer simpler replacement sentences, or a simpler *recipe* for matching them with ordinary claims? Finally, how do recipes of this sort relate to the theories of natural language sought by semanticists? Answering these questions is crucial if we want to get clear on how best to evaluate reductive theories of reality that make use of the ‘paraphrase strategy’.

1. Avoiding tornados, saving appearances

Some metaphysicians think the world can be described completely without mentioning tornados. They think it's better, from a metaphysical standpoint, to talk about destructively whirling air particles. But they would still run for cover if you yelled ‘A tornado is coming!’—at least, if they trusted you. And they wouldn't fault you for not using a metaphysically kosher mouthful when you warned them. No time for ‘carving nature at the joints’ when a tornado is coming!

Other metaphysicians embrace talk about tornados. Some posit an abundance of objects built up from every way of combining time-slices of fundamental particles. On this view, tornados can be identified with mereological sums of ‘temporal parts’—sums that we happen to take a special interest in. But the same cannot plausibly be said when it comes to various other categories of thing, such as flaws, dearths, sakes, rainbows, stocks and

¹ [Acknowledgments omitted for blind review]

bonds. Most metaphysicians will want to resist including *these* things in their official terminology— however useful it is to talk about them in every day life.

This impulse to reduce or eliminate a category of entity stems from two convictions. First, whatever facts are conveyed by ordinary sentences that make mention of entities in that category, they can be *accounted for* or in sentences that do not.² And second, the latter sentences are *metaphysically better* than the originals. Together, these two ideas rule out some alternative meta-metaphysical theses. For example, they rule out *Aristotelianism*, according to which a theory of reality should say how tornados are grounded in other things—a task that requires mentioning tornados.³ And they also rule out *deflationism*, which denies that there is any sense in which the metaphysician’s way of talking about the world is better than the ordinary one.

Let us grant that, for metaphysical purposes, there is something non-ideal about the way that ordinary tornado-talk describes the world. There is still room for disagreement about *why* this is. According to the *strong eliminator*, an utterance of ‘A tornado is coming’ expresses a falsehood, even if it can be used to get across various truths. Meanwhile the *weak eliminator* denies that anything false is expressed by such an utterance at all—at least in the right context. But in such a context, the sentence is being used in a ‘loose and misleading’ way, one not suited to metaphysical theory-building.⁴ Speaking strictly, there are no tornados.

Reducers are even more conciliatory. They hold that there is nothing loose or misleading about ordinary utterances of ‘A tornado is coming’—they can be straightforwardly true, even when uttered with all the strictness and seriousness that a speaker can muster. But a metaphysical description of world should not contain such a sentence, because the fact that a tornado is coming can be fully accounted for by a description about how things are with air-particles. And such a description would be more *metaphysically perspicuous* than one that mentioned tornados—its terms would be more *natural* or *joint-carving*, and the structure of its sentences would better match the structure of the facts they represent.⁵ Moreover, since we need the terminology necessary to describe the

² Section 3 below concerns what exactly it takes to ‘account for a fact’ in this sense.

³ See e.g. Schaffer (2009), Fine (2009); for some objections see Wilson (n.d.).

⁴ Van Inwagen (1990, 108-114) compares such sentences to ‘loose and misleading’ talk exemplified by the ordinary claim that ‘the sun moved behind the elms’: on his view, the ‘strict and philosophical’ sense of that sentence is false. (This approach is consistent with various claims about the ‘literal meaning’ of the sentence as uttered at that context.)

⁵ The notion of naturalness is set out in Lewis (1983, 367). For an extended discussion, see (Sider 2011), chapters 1 and 2

movements of air particles *anyway*, adding ‘tornado’ to the vocabulary of our theory of reality would be ideologically extravagant.

Reducers face some further choice-points. One pressing question is to what degree one’s theoretical language should be thought of as a regimented form of any natural language—say, English. Presumably the reducer wants to avoid the ambiguity and context-dependence of English, as well as any insufficiently joint-carving expressions. She could trim down her stock of expressions to avoid these things, while also adding some theoretical terminology. But there are some difficult questions that arise when it comes to her choice of *quantifiers*.

The reducer must decide whether the *quantifiers* of her theoretical language—assuming it employs some—have the same meaning as any quantifier expressions of natural language. Some potential trouble attends the idea that they do have the same meaning. To start with: won’t her theoretical quantifiers still range over tornados? After all, she has just granted that the English sentence ‘There are tornados’ is straightforwardly true. This appears to entail that the ordinary English existential quantifier ranges over tornados, so adopting that quantifier in one’s theoretical language means adopting a quantifier whose domain contains tornados. One response for the reducer is to acknowledge this result but deny that her theory is committed to everything the quantifier ranges over. She might vie instead for a criterion on which a theory only accrues ontological commitments to the *minimal* domain of entities required to make the quantificational claims of that theory true—that is, a theory is committed to those entities that would have to exist in order to satisfy the corresponding open sentences. Since her theory has no quantified sentences that require *tornados* to exist per se (sentences like ‘There are tornados’), it involves no commitment to tornados.⁶

The more popular alternative is to accept a form of *quantifier variance*, the thesis that an expression playing the linguistic and inferential role of a quantifier in one language need not have the same meaning as an expression

⁶ Notice that the theory in question does not contain a one-place predicate like ‘tornado’, so it will not say ‘There are no tornados’ either. Arguably, after making positive claims that require for their truth that there are just so many (say) particles and regions and sets and so on, our theory needs some analog of the claim that absolutely nothing *else* exists. Otherwise it cannot be said to have accounted for all the facts about the world completely—it has not, inter alia, ruled out ghosts. But any unrestricted claim to the effect that nothing else exists will be false. Here our theorist may claim that she needs no such negative claim (adopting a particularly weak version of the ‘D-project’— see §3), or include a sentence expressing a primitive Totality relation (Armstrong 2004, 72), or hold that negative sentences of this sort get to be ‘true by default’ since their prejacent lack anything to make them true (Mellor 2003: 214; Simons 2008: 14).

playing that role in every other possible language.⁷ Unlike deflationists, of course, a reducer will not be *egalitarian* about possible meanings of expressions that play the the quantifier role. In particular, she will consider the quantifier-like expressions of her theoretical language to be better suited for describing reality than those employed in ordinary talk.⁸ (This kind of language is sometimes called ‘Ontologese’.)

For those who find inegalitarian quantifier variance plausible, reduction is a better approach than elimination. If we can interpret the English quantifier so that ordinary sentences positing tornados are straightforwardly true, we needn’t join the strong eliminator in denying what seem to be obvious facts about the weather, or in claiming that people routinely utter falsehoods with complete sincerity in order to convey truths. And we needn’t join the weak eliminator in classifying a wide swath of language use as ‘loose and misleading’, when it seems just as strict and serious as any other talk we engage in.⁹ After all, it is a matter of *use*, not a matter to be settled by metaphysics, whether the folk are using quantifiers ‘loosely’ when they talk about tornados; so the weak eliminator takes on a significant burden for a semantic theory.¹⁰

(Quantifier variance also raises a concern for the Aristotelian, who holds that existence questions usually have a trivial answer of ‘yes’. This allows her to treat Moorean claims about what exists at face value; the task of ontology is not to say what exists so much as it is to say what grounds what, and how. But the Aristotelian does face *some* question about what exists. Does she go so far as to embrace weird mereological fusions, and things with bizarre modal profiles like snowdiscalls and in-cars?¹¹ If she accepts a kind of mad-dog realism according to which there is an object for every modal profile that can be associated with

⁷ See Sider (2011, §9.5-9.6) for a discussion of some pitfalls that accompany attempts to more rigorously define the idea.

⁸ See Sider (2009, 2011, ch. 9); Dorr (2005).

⁹ Is there anything that the folk regularly quantify over in which they speak ‘strictly’ as opposed to ‘loosely’? If so—what is the difference in their patterns of use? If not— how can such a pervasive expression come to have a ‘strict’ use that is rarely employed?

¹⁰ In contrast, apparent quantification over *sakes* is semantically non-standard. ‘Sake’ acts like a noun in some idiomatic contexts but it can’t freely be modified and always has a possessive attached. In fact, ‘Sakes exist’ is arguably ungrammatical.

¹¹ *Snowdiscalls* are like snowballs but can survive being squashed into pancake-shape (Sosa (1993). *In-cars* overlap with cars inside garages but go out of existence when those cars leave their garages (Hirsch 1982). Even stranger are things that not only have strange modal profiles, but are also mereologically composed of scattered and unrelated parts—for example, an object that is made up of Trump’s toupee and Palin’s glasses and goes in or out of existence every time I sneeze.

every fusion of temporal parts, this would mitigate any Moorean grounds for her view. There are negative Moorean facts—for example, no nose-shaped fleshy thing on my face goes out of existence when I giggle. If the Aristotelian quantifier-variantist rejects mad-dog realism, the question arises why a more expansive language that did quantify over such objects wouldn't be better for doing metaphysics—since in such a language one could articulate what grounds things like snowdiscalls, incars, and giggle-noses!)¹²

The purpose of this paper, however, is not to argue in favor of reduction over elimination, but to discuss the project of 'saving the appearances' that is critical to both approaches. I will proceed as though ordinary claims about tornados are straightforwardly true. But eliminators also need to account for the truths they take to be conveyed by tornado-sentences of ordinary language. For this reason, much of what I say will apply, *mutatis mutandis*, to their approach as well.

2. More than Moore

The guiding aim of metaphysics is to build a theory of reality—a 'total theory', as Lewis puts it, 'the whole of what we take to be true'.¹³ The ideal theory of reality will therefore be *complete* in the sense that it accounts for every fact; and the rough outline of such a theory that a metaphysician actually offers must be *completable in principle*. If a nascent theory simply cannot be fleshed out in such a way that it accounts for all the facts, it can only ever be a theory of some *part* of reality. (Thus *indispensability arguments* attempt to show that no complete theory can avoid commitment to a certain kind of entity.)

If one wants to challenge whether a theory can account for all the facts, one might as well start with the most obvious ones—*Moorean* facts. Thus if one's theory does not mention tornados, it should still somehow account for the fact

¹² A natural reply for the Aristotelian, aside from denying quantifier variance, is to say that in a language whose quantifiers have different meanings than they do in English, expressions like 'grounds' will *also* have different meanings. They express grounding* relations, rather than grounding relations. Thus none of these other languages are suited for the question we're interested in, namely what grounds what. One problem with this answer is that there does not appear to be anything *objectively* interesting about this question—especially given that it is wrapped up with a quantifier meaning that does not in any way appear to be particularly joint-carving. This brings the Aristotelian uncomfortably close to the deflationist.

¹³ 'That which is our professional concern', he says, is 'total theory: the whole of what we take to be true' (1993: 3).

that a tornado is coming.¹⁴ The hope is that one's tornado-free theory can contain a sentence—perhaps one about the movements of air particles—that somehow accounts for this fact. Saying anything very specific about such a sentence would require, of course, a view about what sort of behavior on the part of air particles would be required to make it the case that there is a tornado nearby. But it would *not* require a view about what makes it the case that the ordinary sentence 'There is a tornado nearby' is true.¹⁵ After all, the ordinary sentence is true partly because it has the meaning it does—but an account of *that* fact would take us far afield from air particles.¹⁶ What's at issue is the part of her theory that accounts for tornados, not the part that accounts for the semantic properties of sentences.

This last point bears some emphasis. The Moorean challenge is sometimes described as demanding 'adequate paraphrases' for various ordinary English sentences, or even providing an account of what makes those sentences true. But paradigmatic Moorean facts are not about language at all—they are about the existence of perceptually obvious things like hands. And indeed, the task of a theory of reality has no more to do with ordinary language *per se* than it has to do with any of the other complex and marvelous phenomena that the world exhibits—such as weather patterns, DNA, and economic systems. A theory of reality-as-a-whole must in principle be capable of accounting for *all* of these phenomena. (Of course, when looking for a way to account for the fact that there is a tornado nearby, it can be natural to semantically ascend and ask oneself what it would take for 'There is a tornado nearby' to be true. But this

¹⁴ For those interested in applying the themes of this paper to eliminativism, replace instances of the schema 'account for the fact that S' with 'account for the fact(s) conveyed by 'S'/expressed by 'S' in a loose context'. And replace instances of "'S' is true' with "'S' conveys a truth/expresses a truth in an ordinary context'.

¹⁵ Of course, it is possible to ask what it would take for there to be a tornado nearby by asking, 'What would it take for "There is a tornado nearby" to be true?' We need only treat it as common ground that the truth of 'There is a tornado nearby' requires only that there be a tornado nearby—in other words, we need only set aside any need for an account of that semantic fact.

¹⁶If her theory is highly reductive, this would involve giving a reductive account of sentences and of meanings, assuming she agrees with Fodor that 'intentionality doesn't go that deep' (1987:97). That is, it would require her to articulate, in the joint-carving terms of fundamental metaphysics, what configurations of reality underlie a situation that counts as involving a true utterances of the relevant sentence.

heuristic question need not be answered *in the language of the theory*—at least until the theory’s ability to account for semantic facts is itself at issue.)¹⁷

The ability in principle to account for every fact is an *absolute* requirement on a theory of reality—not a virtue to be weighed against others in a comparison with rival theories. For this reason, Moorean arguments have a special place in the standard methodology among those who think that disputes in contemporary metaphysics are both substantive and tractable. The broad working consensus is that competing theories are evaluated using a set of theoretical desiderata like simplicity and explanatory power, in much the way we evaluate empirically equivalent scientific theories. Following Ted Sider, call this the *quasi-scientific method* or QSM. The idea is to

treat competing positions as tentative hypotheses about the world, and assess them with a loose battery of criteria for theory choice. Match with ordinary usage and belief sometimes plays a role in this assessment, but typically not a dominant one. (Sider 2009: 385)

A paradigm of that method is David Lewis’s exposition of the theoretical benefits of his plurality-of-worlds hypothesis, which he took to provide ‘a reason to think that it is true’ (1990:3).¹⁸

So where do Moorean arguments fit in? The QSM involves comparing theories by measuring them using several graded criteria, and then performing a cost-benefit analysis that weights each criterion in comparison with the others. But Moorean arguments concern an absolute requirement on a theory of reality, rather than a virtue to be weighed against others. If a theory of reality cannot be fleshed out in a way that accounts for some fact about reality—if it cannot consistently be made complete—the jig is up. The analog for scientific theories is compatibility with available data: gradable criteria of theory-choice are relevant only when comparing theories that pass the initial test of adequacy to the known facts.

¹⁷ Semantic ascent can be especially tempting when we are wary of semantic externalism: we may take ourselves to be entertaining an epistemic possibility which, if true, might well shift the meanings of some key expressions from what we take to be their actual meanings. In that case I think the question at issue can still be asked most perspicuously in the material mode: considering as actual a world where things are as this theory of reality says they are, what would it take—in terms of the language of the theory—for there to be a tornado nearby?

¹⁸ ‘We have only to believe in the vast realm of possibilities, and... we find the wherewithal to... improve the unity and economy of the theory that is our professional concern... The [theoretical] benefits are worth their ontological cost’ (Lewis 1990:4).

Despite this difference, gradation is relevant to Moorean arguments—in the form of degrees of doxastic confidence. To begin with, it is sometimes difficult to *prove* that a theory cannot account for a given Moorean fact. There may be reason to doubt that it could, but nothing definitive. In addition, the relevant fact itself need not be something known with *certainty* to play a role in an argument challenging the completability of a theory. Any positive degree of belief that p can furnish evidence against a metaphysical theory on which p turns out false. Since it would be odd to use ‘Moorean’ as a label for arguments based on middling credences, I will use the expression ‘saving the appearances’ for this sort of consideration. A theory’s success in saving the appearances must be relativized to an individual, and measured by weighting each vindicated belief by the strength of her credence in it.

In short, completability is all-or-nothing—but it can be hard to know whether it applies to a given theory. In contrast, the QSM’s criteria are gradable. And knowing to what degree they apply to a given theory is relatively easy—the hard question is how to weigh them.

3. Accounting-for

Can we clarify what it is for a reductive theory to *account for* a fact?

There are, broadly speaking, two options, corresponding to two visions of the task of metaphysics. Kit Fine describes them this way:

The E-project is concerned with *saying* what can be said in the most fundamental terms, while the D-project is concerned with *describing* what can be described in the most fundamental terms. We can easily bring out the difference between the two projects with the case of disjunction. I can say ‘p or q’ and it is not clear that this can be said except by using disjunction or the like. But suppose now that I correctly describe the world by means of the sentence ‘p or q’. Then the use of ‘or’ is dispensable, since I can alternatively describe the world by means of p or q, depending upon which is true. Thus even though ‘or’, or the like, may be indispensable for saying what we can say, it would not appear to be indispensable for describing what we can describe.

To put things a little differently, if our task is the E-project, adequately accounting for a fact requires saying *what it is* for that fact to be the case. On such a view, ‘There are particles whirling’ accounts the fact that there are tornados iff: for there to be a tornado just is for there to be air-particles

whirling.¹⁹ (Or: there is *nothing more* to there being a tornado than that there are air-particles whirling.)

But if our task is the D-project, things are different.²⁰ On this view, accounting for a fact only requires saying *what makes it the case*— what underlies that fact. For example, we do not want to say that for there to be tornados or ghosts is for air particles to be whirling vortex-wise. Still, we aren't missing out on a complete description of the world if we simply say what actually *makes it the case* that there are tornados or ghosts. 'There are air particles whirling' accounts for the fact that there are tornados or ghosts all by itself, because what makes it the case that there are tornados or ghosts is that there are air particles whirling.²¹

(Note that the E-project's central expression, 'just is for', can combine more than two infinitivized sentences. For example, it is plausible that for *a* and *b* to be red just is for *a* to be red and for *b* to be red.²² Correspondingly, we should allow that one way for a fact to be accounted for by a theory is for the latter to contain multiple sentences that *together* account for that fact. Likewise for the D-project's central expression, 'what makes it the case that':

¹⁹ More generally, let 'S' and 'S*' be schematic symbols such that instances of the schema have a sentence in place of 'S' and a full infinitivization of that sentence in place of 'S*'; and likewise for 'R' and 'R*'. Then: 'S' accounts for the fact that R iff for R* just is for S*.

²⁰ See Fine (2012), Rosen (2010), Melia (2005). The difference between E-theorists and D-theorists is mirrored in the literature on truth-makers—truth-maker *maximalists* claim that all truths need truth-makers, while truth-maker optimists disagree. (See e.g. Mellor 2009, Cameron 2008, Merricks 2007, Bigelow 1988.) The kind of D-theory I have in mind shouldn't be taken to involve commitment either to propositions or to truth-maker.

²¹ 'Just is for' and 'what makes it the case that' are related to the notion of *metaphysical analysis*: see Dorr (2005), Sider (2011 §7.4). But the latter expression can be taken to suggest that what's at issue is a relation between two sentences: an ordinary sentence and its theoretical replacement. But if our goal is to say what it is for a theoretical sentence to account for a given fact, how did *pairs* of sentences enter into the equation? Completeness can't just be a matter of providing a metaphysical analysis for every true sentence of some natural language. Even facts that can't be expressed in a natural language—if there are any—must be accounted for.

²² The connective is therefore 'weak' in the sense of Fine (2012).

what makes it the case that there is a tornado and an electron is that there are air particles whirling and that there is an an electron.²³)

It remains to decide whether the E-project or the D-project best characterizes the task of metaphysics! Much hinges on the right answer.²⁴ Here is one consideration that may seem to favor the D-project over the E-project, as a characterization of the task of metaphysics. Suppose that we follow Sider in holding that the all of the *expressions* used by a theory of reality—not just its predicates—should be maximally joint-carving. Now, assume that the fact that p or q can be accounted for by the sentence ‘ $\sim(\sim p$ and $\sim q)$ ’. In that case a theory of reality needn’t contain *both* logical connectives. But could it really be the case that reality has a structure best matched by ‘and’ as opposed to ‘or’—or vice versa?²⁵ This choice can be adopted by the D-project, which arguably renders both disjunction and conjunction unnecessary. (Disjunctive facts are handled by accounting for a single disjunct; and conjunctive facts can be accounted for by multiple sentences collectively.) There are, of course, other options for the E-theorist. One is to accept that the structure of reality can be equally well matched by two different languages. Another is to simply retract the idea that joint-carvingness properly applies to logical expressions.²⁶

²³ A similar repudiation of the E-project is implicit in a recent idea from Rayo (2008) and Williams (2011). They employ an ‘in reality’ operator to articulate how things must stand with a theory’s canonical objects in order for various facts to obtain; but they also allow quantification into the scope of such an operator, as in ‘the fact that there is a tornado requires that, for some xs that are tornado-parts: in reality the xs are moving thusly’. Instead of providing a canonical sentence that accounts for the fact that there is a tornado, this formulation specifies from outside the language of the theory, how things must stand with the objects it postulates. I will set aside this intriguing approach in what follows.

²⁴ For example—what if a Moorean or indispensability argument establishes that a theory is not E-completable? This is a fatal objection on one picture of the task of metaphysics, and on the other it is no objection at all.

²⁵ See Sider (2011 §10.2) One might be tempted to give up negation on simplicity grounds and use only NAND (\uparrow); but then why not NOR (\downarrow) instead?

²⁶ Does this threaten the idea of inegalitarian quantifier variance? Not obviously: as Sider argues, a shift in quantifier meaning plausibly shifts the meanings of every other expression in a language. Perhaps all the relevant inequalities between languages can be measured in terms of the relative joint-carvingness of their predicate meanings.

4. Saving the appearances... in principle

It turns out that actually *providing* a sentence that accounts for a given fact about tornados is more difficult than one might have hoped. (And tornados are a fairly easy case!) This gives rise to several questions: (1) suppose that accounting for a tornado-fact appears possible *in principle*, but is highly impractical or even unachievable given human limitations. Does this provide us with at least a *ceteris paribus* reason to reject it? (2) In general, where does the burden of proof lie when there is a dispute about whether a reductive theory can account for some fact?

Analogous questions arise in the case of fundamental physics. In principle, the best physical theory should be capable of accounting for all the facts in its domain, including facts about why my coffee mug doesn't just float off when I set my mug down. So applying the best fundamental physical theory to the case of my coffee mug should be possible *in principle*. But no physicist would bother actually providing a specific account of the microphysical substructure underlying my cup and the table. Instead, physicists focus their energies on offering the most explanatory contours of the complete physical theory.²⁷ This does not mean my coffee mug is *irrelevant*—it would be worth special attention if it somehow threatened to embody a counterexample to the theory. But barring any reason to suspect that, it's sufficient to have a general idea of the sort of sentences a complete physical account of my coffee mug might have.

Similarly, a metaphysician needn't actually *provide* a sentence that accounts for an obvious fact every time a challenger asks her to. It is enough to make a reasonable case that such a sentence is available in principle. Any merely *practical* barriers to articulating the relevant sentence should be considered irrelevant from an *evidential* point of view. For highly reductive theories, one such practical barrier will be mind-bending complexity. (Even a theory with a very simple axiomatic base will logically entail infinitely many highly complex sentences— and some of those sentences may account for Moorean facts.) But such practical concerns do not amount to a persuasive Moorean argument.

Consider an example discussed by Joseph Melia (1995). Suppose we know that the average star has two planets; but we don't want to quantify over average stars. Is it inescapable that in accounting for the fact that the average star has two planets, we must appeal to numbers—as in 'The number of planets divided by the number of stars is two'? Melia replies, in effect, by rejecting the implicit assumption that our theory must be E-complete. He points out that a theory of reality can simply say, as it may be, 'There are one

²⁷ When detailed descriptions of actual goings-on in space-time are offered, they are carefully chosen to confirm or disconfirm important theories.

septillion stars and two septillion planets’. That sentence does not quantify over numbers.²⁸ But suppose our goal is E-completeness. In that case, why not point towards the possible theory that accounts for the fact that the average star has two planets with the sentence:

1. There is one star and two planets, or there are two stars and four planets, or there are three stars and six planets...

This time our inability to express the theory is not due to a lack of empirical knowledge, but due to a lack of paper, ink, and time. But surely our cognitive and expressive limitations could only give us a reason to *proceed as though* numbers exist, as opposed to a reason to *believe* they do. It would be very odd if our limitations afforded us reasons of the latter sort, reasons that would in principle be unavailable to a deity whose language allowed infinitely long disjunctions.²⁹

In short, the need to ‘save the appearances’ applies only to the sort of ideal finished theory towards which a flesh-and-blood metaphysician can only gesture. An effective indispensability argument must argue that there is some obvious fact that no theory without entities of kind **K** could even in principle account for. But this can be a difficult case to make, especially if the reductive theory has a fairly rich vocabulary and is stated in an expressive language. As a result, this dialectical situation is likely to result in frequent stalemates. On the one hand it will often be far from clear how to make a case that some fact (say, that there are donkeys) cannot even in principle be accounted for by a certain kind of reductive theory. On the other hand, it may be equally difficult to make a case that it *can*.

²⁸ We could likewise avoid the numerical expressions entirely using negation, identity, and two septillion variable expressions. (Just as ‘There are two planets’ can be replaced with ‘ $\exists x \exists y (Px \& Py \& (\neg x=y) \& \neg \exists z (Pz \& z=y \& z=x))$ ’.)

²⁹ [Reference to Author’s work omitted.] It follows that we needn’t accept the ontological commitments of *our* best theory, in the sense of a theory actually expressible by us—at best we have to accept what we know would be the ontological commitments of the best possible theory. This kind of reply to the indispensability argument will carry no weight with pragmatists like Quine, for whom answering ontological questions involves choosing the most convenient framework rather than discovering a matter of fact. But metaphysical realists view the QSM as a way to get evidence about which theory best matches reality. The fact that we cannot fully articulate the relevant sentence can only generate pragmatic indispensability arguments of a sort that will fail to convince a metaphysical realist.

5. Saving the appearances and interpretive charity

Another area of potentially intractable disagreement concerns the question of whether a sentence actually accounts for a given fact—even assuming the truth of a given theory of reality.

Here is an example. According to Bishop Berkeley's theory of reality, there are spirits and the ideas had by those spirits—and nothing else. But Berkeley goes on to claim that for there to be a table is for there to be certain ideas bundled together in spirits, for God to be disposed to continue to produce these ideas, etc. So if he were to flesh out his theory, he could account for ordinary facts about tables and chairs, using sentences about various perceptual ideas bundled together in spirits. Now contrast Schmerkeley, another idealist who differs from Berkeley only in that he denies that there are objects like tables and chairs; he holds that for there to be such things would require a mind-independent spatiotemporal reality. In short, Berkeley and Schmerkeley share a theory of reality. They differ only on whether any sentences of their theory actually account for facts about tables and chairs.

Notably, only Berkeley offers to save the appearances. But this does not by itself seem like a reason to prefer his view over Schmerkeley's. Whether appearances can be saved is a reason to prefer one *theory of reality* over another. But the issue between Berkeley and Schmerkeley is whether—assuming the same theory of reality as actual—what it is for there to be a table is for there to be perceptual ideas of a certain sort bundled together in spirits. And one might well hold—indeed, I am tempted to think—that *Schmerkeley* is right on this question. But it is hard to say exactly why that is. I suppose it has to do with a sense that, considering their shared theory of reality as actual, I intuit that our perceptual states elicit false contents as of an external word: we are systematically deceived by the perceptual ideas God gives us. If this is right, then there are Moorean reasons to reject this kind of idealism: *contra* Berkeley, it does not account for the fact that I have hands.

In general, the success of Moorean arguments will hinge on whether the relevant appearances actually get saved, given the theory under attack. To take another example: let us suppose that as a matter of fact there are many concrete universes spatiotemporally unrelated to ours, and that reality contains no other ground for modal talk. Given these assumptions, is it plausible that what it is for it to be possible that donkeys talk is for there to be talking donkeys in one of these discrete universes? If not, Lewis's modal realism is open not only to the charge of ontological profligacy, but also the charge that sentences about distant concrete universes would not account for the modal

facts he alleges they do—even assuming such universes exist.³⁰ The second objection is fundamentally Moorean: Lewis is wrong about which facts would be accounted for by his theory of reality if it were true.

Unfortunately there is no agreed-upon methodology to identify which facts a theory of reality can account for. Everyone seems to agree that *interpretive charity* is relevant—but is this a *ceteris paribus* inclination to preserve the *truth* of what people say, or the *reasonableness* of their beliefs? Berkeley’s view would benefit from the former approach, but arguably not the latter. Schmerkeley can say: ‘On my view people have misleading perceptions as of there being a mind-independent external world. So it’s perfectly reasonable and expected for them to form false beliefs about whether there are tables.’ By my lights, this reasoning does undermine any charitable pressure in favor of Berkeley’s view.

It is crucial to distinguish the criterion of interpretive charity from that of saving appearances. A Moorean objects to Schmerkeley’s view by treating ‘I have hands!’ as a *premise*. That is something entirely different from saying: ‘The proportion of ordinary sentences that your view deems true is quite low, as compared with mine—a fact that we should take into consideration in our overall assessment of the two theories.’ The Moorean makes a direct claim about the world that must be upheld, rather than appealing to some theoretical benefit having to do with semantic interpretation.³¹

The two criteria also have sharply different methodological roles. Interpretive charity is relevant for assessing which putative facts are actually accounted for by a theory of reality; while saving the appearances is relevant for choosing among theories once this assessment has been made. Suppose I am deciding between theories of reality in part based on whether they can account for the fact that I could have been taller. Before applying this consideration—an instance of saving the appearances—I must first look to interpretive charity

³⁰ This is similar in spirit to Plantinga’s (1987) objection to Lewis’s modal realism.

³¹ Even charity-to-reasonableness will clearly need refinement. For example, more weight should be placed on truths that are treated by speakers as conceptual or linguistic truths, and less weight on truths that are taken to be empirical. To illustrate: suppose reality doesn’t contain any supernatural entities of any kind, and that it’s downright irrational to believe in any. Now suppose we’re interpreting the language of some community that talks as though they believe in God. Someone might suggest interpreting God talk as being about the fact that there is love and beauty in the world, in order to ensure that at least some sentences and beliefs of the community (like ‘God exists’) come out true and rational. What makes this a crazy interpretation is that people treat it as analytic that God has powers of various kinds, is a person, and so on. And preserving the truth/rationality of these connections, enshrined in claims like ‘If God exists, God is a person’, trumps preserving the truth/rationality of other claims they make.

(among other things) to decide which theories actually account for the putative fact in question. The latter assessment prescind from any antecedent attachment I might have to the putative fact.

Why, it is worth asking, do we defer to interpretive charity at all? For reasons just given, it cannot be driven by the fact that we antecedently take the relevant ordinary sentences to be true. (Thinking that ‘I have hands’ is true is not a reason to hold that idealism accounts for that fact— it’s a reason to reject idealism if it doesn’t!) Nor does interpretive charity seem to be a fundamental epistemic inclination, like (perhaps) induction. Instead, arguably, our bias towards matching ordinary truisms with true (or reasonable) replacement sentences for ordinary truisms is justified by the fact that the practice of predicting and explaining behavior by attributing beliefs and desires (i.e. the ‘intentional stance’) would not work nearly so well if the beliefs we attributed to each other were false and unreasonable.³² The practical success of folk psychology will be much easier to explain if the bulk of ordinary beliefs are pretty reasonable.

If this is our motive for charity, it is not a *sui generis* virtue. An appeal to charity amounts to a prediction—given certain background assumptions—that a rival sketch of reality will incur explanatory costs when it is extended to account for the success of folk psychology. But, at least in principle, that prediction might not bear itself out. Suppose there is some acceptably simple theory of reality which, on a natural way of matching replacement sentences to ordinary sentences, can account for the success of folk psychology *without* many ordinary beliefs coming out true/reasonable. (This is so, perhaps, for an idealism involving a God who doesn’t mind deluding created spirits: at least when it comes to the *truth* of ordinary beliefs.) If I am right, this would not give us reason to prefer another way of interpreting ordinary claims on that theory of reality. If anything, it might give us a reason to reject that theory of reality, because (e.g.) *I do have hands!*—not for reasons of interpretive charity.

6. Varieties of simplicity

Simplicity or parsimony is perhaps the central theoretical virtue by which the QSM would have us evaluate theories of reality. But there are several axes along which the simplicity of a theory can be measured, and there is plenty of disagreement about how they should be weighted.

It is standard to treat simplicity as coming in two varieties: *ideological* and *ontological*. The first has to do with the simplicity of a theory’s language, and the

³² See (Lewis, 1974).

second with the number of entities it is ‘ontologically committed’ to. Consider, for example, three simple but D-complete theories: ‘Fa’, ‘Fa, Ga’, ‘Fa, Fb’.³³ The difference in simplicity between the first two theories would traditionally be considered ideological, whereas the difference between the first and the third would be considered ontological. But it is not as though the one difference has only to do with the *language* of the theories, while the other also has to do with the *world* as they describe it. The second and third theories *both* require the world to be more complex than the first theory does, as well as using a more complex language than the first theory. For this reason, it’s more useful to distinguish between a theory’s ideological simplicity and its *worldly* simplicity, where the latter goes beyond counting how many things there are according to a theory. It also has to do with how *simple* those things are according to a theory.³⁴

How then should we measure ideological simplicity? Simply adding up the number of primitive expressions in a theory can be misleading if those expressions are is sufficiently ‘gruesome’ or non-joint-carving. Imagine that God defines ‘hunky dory’ as a predicate expressing exactly how things actually are in general: still, ‘Things are hunky dory’ as a theory of reality may be simple, but it is far too gruesome to be taken seriously. In fact—arguably what *makes* this sentence gruesome is precisely the contrast between its simplicity and the complexity of the reality it describes. If this is the case, its simplicity is not a virtue *outweighed* by its gruesomeness: its simplicity is a *vice* given that the reality it describes is so complex.

This line of reasoning suggests that ideological—that is, grammatical—simplicity should not be thought of as a virtue to be weighed against worldly simplicity. If we have—as I think we do—an independent grip on how simple the world is according to a theory, we should replace the alleged virtue of a theory’s linguistic simplicity with the virtue of *matching* the simplicity of the reality it describes. A theory should be simple when describing simple things—for example, a fundamental particle’s having a fundamental property—but also rich enough to mirror the tangled detail of reality’s messy bits. The best theory will have a language that is exactly as complex as it needs to be and no more.

³³ I’m assuming the expressions of these theories are all natural and non-redundant.

³⁴ Melia contrasts the question whether a theory can more simply express the facts than another, with the question whether ‘the *world* is a simpler place’ according to it. ‘The simplicity I value attaches to the kind of world postulated by the theory— not to the *formulation of the theory itself* (2000: 473). To my mind it is also a virtue of the formulation if its simplicity reflects the simplicity of the world it describes.

In sum: when employing the QSM, favor theories according to which the world is a simpler place overall. And secondarily, favor theories whose linguistic structure matches the structure of the reality they describe.

(As an aside, it is worth asking after the relationship between fact-simplicity and joint-carvingness. Can one be understood in terms of the other? If we take the notion of the naturalness of expressions as primitive, we could define fact-simplicity in terms of the number of expressions required to express the relevant fact in a maximally joint-carving language. But it seems less than ideal for a metaphysical theory to have a primitive notion applying to expressions: after all, linguistic items are not likely to be found among the ground-floor furniture of the world.³⁵)

Another option is this. We can recover much that is useful about joint-carvingness if we take as primitive a notion of relative fact-simplicity instead. For example, for a to be blue is simpler than for a to be grue. (Officially this can be expressed with a primitive sentential connective: it is simpler that ϕ than that ψ .) The problem with ‘grue’ is that its lack of structure belies the highly complex facts that are expressed by even atomic sentences containing it.³⁶ Roughly, on this view, the most joint-carving expressions will be those that can more simply described simpler facts. Thus, for example, the fact that two particles are charged is simpler than the fact that one is charged and the other is massive. So a theory should use fewer expression-types to articulate the first fact—in a standard first-order language, a single predicate rather than two.³⁷ On this view, similarity is a kind of simplicity.)

³⁵ Proponents of naturalness have sometimes preferred to apply that notion in the first instance to *properties*, defining natural predicates accordingly. But for those like Sider who hope to take the relevant notion ‘beyond the predicate’, it is trickier to analyze the metalinguistic feature in terms of a primitive feature of reality. One approach is to take as primitive the notion of *natural facts*, and treat natural terms as those that make for syntactically simple expression of those facts. (Sider himself opts for a more subtle strategy involving a primitive operator that applies to ‘notions’: 2011, §6.)

³⁶ Well, this is also a problem with ‘blue’— but ‘grue’ is worse!

³⁷ Does this give us any reason to prefer one quantifier over another—assuming we need one? I am inclined to think that for *something* to be massive is simpler than for *everything* to be massive—even if there happens to be just one object. (The universal fact puts constraints on the world as a whole.) On the present view, this would be a reason to take the existential quantifier as primitive rather than the universal.

7. The simplicity of ordinary facts

There are other considerations of simplicity that can be used to evaluate theory of reality, even aside from ideological and worldly simplicity.

First— consider what we might call ‘ordinary-fact simplicity’. Suppose we are comparing two theories of reality, T1 and T2, that score equally well on the measures of simplicity just set forth. But suppose they differ in the following respect: when it comes to the various boring facts that comprise the ‘appearances’ we typically talk about, the sentences of T1 that account for such facts are relatively simple; while the sentences of T2 that account for such facts are much more complex. Is this a reason to prefer T1 over T2? I will argue that it is not.³⁸

There are, of course, *pragmatic* reasons to like T1 better than T2. After all, T1 will be easier to use and easier to grasp with our native conceptual apparatus, since it accounts for facts we tend to think about using fairly simple sentences. (Keep in mind that the two theories are tied as a matter of overall simplicity; so this difference must be balanced out by the fact that T2 contains simpler sentences *elsewhere*—perhaps in a part of the theory that accounts for facts we cannot or would not express in natural language.) But does this practical advantage give us any *evidential* reason to choose T1 over T2? This amounts to asking whether we have any prior reason to expect that the sorts of facts we talk about from day to day are especially easy to articulate in a joint-carving language. But why should we think that? Considering the bafflingly intricate neural structures that realize mental representations, the messy complexity of human interests, and the tinkering evolutionary origins of both, it would hardly be surprising if the sorts of facts that we typically concern ourselves with are *objectively* incredibly complex, and therefore prohibitively difficult to articulate in a genuinely joint-carving language. We should first settle on a theory of reality, and then let *it* tell *us* how metaphysically complex the ‘world of appearances’ is.

By analogy, imagine proposing to physicists that we should accept an otherwise inferior fundamental physical theory P1 against the standard theory P2, on the grounds that P1 considers medium-sized dry goods like tables and trees to relatively simple physical phenomena, whereas P2 treats them as comparatively complex. (Suppose that both theories entail the existence of highly complex physical phenomena; the difference lies in whether the sorts of things we talk about with simple sentences are among them.) Would it ever be

³⁸ To clarify: we are not here comparing alternate views about which sentences actually account for the relevant facts, assuming a shared theory of reality. For such a comparison the comparative simplicity of the proposed sentences is surely relevant.

worthwhile to sacrifice the economy of a physical theory's explanatory base in order to allow for a simpler articulation, in the fundamental language, of the facts we concern ourselves with in everyday life? It is hard to imagine physicists taking such a proposal seriously. Instead, we should *first* settle on the best physical theory and *only then* make judgments about how physically complex the objects of our day-to-day discourse are.

In reply, here is a tempting argument for preferring T1 over T2.³⁹ Suppose we focus on those facts expressed in ordinary language that play an explanatory role in higher-level theories—for example, in meteorology, social psychology, and the like. Insofar as it is a theoretical virtue for *explanations* to be as simple (and natural) as possible, a theory of reality that can account for facts involved in special-science explanations using simpler sentences will do a better job of vindicating our use of those explanations.

I find this line of reasoning suspect. I do think that in our special-science theorizing we aim to mention, as explanations, the simplest (or most natural) facts that will do the job. But that's a matter for comparing explanations in the special sciences, *given* a theory of reality. It's not a criterion for choosing *one theory of reality over another*. For example, it may well be that our explanations of each others' behavior will be best interpreted as sensitive to the most natural available properties that in fact correlate with the relevant behavior. But none of this gives us reason to suspect that reality will underwrite some minimum objective threshold of naturalness for the facts involved in our psychological explanations. After all, it might be that the relevant *explananda* are objectively much more complex (or non-natural) according to one theory, so it would be unsurprising if the relevant explanations are more complex as well. Indeed, we have been assuming that the two theories are equally simple overall—so *both* theories will presumably have highly complex explanations for explananda that (by their own lights) are highly complex. The difference lies in the fact that one of them treats an aspect of reality that we find particularly interesting as objectively more simple. I know of no way to defend a theoretical bias in favor of this assumption.

8. NL-friendliness

Here is another axis of simplicity along which two theories might differ. Suppose that T3 and T4 are complete and equally simple theories, in all the senses described so far. But they differ in one respect when it comes to facts that are expressible in English. T3 admits of a simpler procedure for identifying which sentence of the theory accounts for the fact expressed by a given English

³⁹ Thanks to [name omitted] for discussion here.

sentence. In other words, the ‘reducing’ sentences of T3 are more systematically related to their corresponding English sentences. For short, T3 is more ‘English-friendly’ than T4.

Is it a theoretical virtue to be English-friendly—or, more generally, natural-language-friendly? Does it give us an *evidential* reason to accept T3? Certainly the fact that T3 is English-friendly makes it easier to be confident that the theory is *complete*—it is easier to check that it can account for a given fact, as long as that fact is expressible in English. But that’s a far cry from the claim that English-friendliness is a virtue of its *own*, a virtue that would bear on theory choice even if we *knew* that both theories were complete.

It is not uncommon to find metaphysicians arguing that some minimal level of NL-friendliness is a requirement for an adequate theory. For instance, Gabriel Uzquiano writes:

[I]f the general strategy of paraphrase is to stand a chance, its proponents must make it plausible that there is a perfectly general and systematic method of paraphrase that will enable one to eliminate the apparent ontological commitment... exhibited by a great range of commonplace statements. (2004)

Uzquiano is not merely saying that a reducer must make it plausible that *there are* sentences of her theory that can account for the facts expressed by commonplace statements. He requires a ‘general and systematic method’—although he does not justify the requirement. Nor does he elaborate on the minimum threshold for systematicity—a feature which, after all, comes in degrees.

One reason one might prefer to have ‘systematic method’ of this type is if one hopes to use it as a *compositional theory of meaning*. For example, David Lewis stressed the usefulness of modal realism ‘for the analysis of language’: it allows for a systematic semantics of ordinary claims directly in terms of items quantified over in the fundamental theory.⁴⁰ And, for Lewis, the ability to provide an elegant and systematic account of the truth-conditions of ordinary sentences makes modal realism particularly ‘serviceable’, which is a ‘reason to think that it is true’.

I will try to call this idea into question, but not because I deny that a theory of meaning should be tractable, systematic, and (at least) fairly compositional. This is an axiom shared by all those who take themselves to be working out a

⁴⁰ Lewis 1986: 40-41. Quantification over properties is understood in terms of sets of possibilities, talk about contents is understood in terms of sets of possible worlds, modal talk is understood in terms of goings-on in other possible worlds, etc.

‘semantic theory’—even though there are several distinct projects that such theorists might be undertaking.⁴¹ For some, the reasons for this axiom are cognitive: the theory must provide a scientifically useful model for how the human language organ systematically combines meanings. A semantics that assigned meanings to complex expressions higgledy-piggledy, without any governing rules that could reasonably be encoded in a human brain, would be a non-starter.⁴² For others, the reasons are instrumental: semantics must systematize language practice in a way that is useful for predicting and explaining linguistic behavior (but needn’t be sensitive to psychological mechanisms that implement that practice).⁴³

If a theory of reality is not very English-friendly, then, its language would not serve very well as the meta-language for a semantic theory of English. But why should we expect to?⁴⁴ For example, as Chomsky writes:

The NP **the flaw in the argument** behaves in all relevant respects in the manner of the truly referential expression **the coat in the closet**—for example, it can be the antecedent of a pronoun and serves as an argument, taking a Theta-role. Suppose now that we make a rather conventional move, and assume that one step in the interpretation of LF is to posit a domain D of individuals that serve as values of variables and as denotata. Among these individuals are specific flaws that can appear in arguments, (cf. “the same flaw appears in both arguments”), John’s lack of talent, and so on.... (Chomsky 1981: 324)

But, he argues, it doesn’t follow that in any deep sense reality must contain lacks of talent: ‘Assigning elements of D is in effect... the construction of [a] level of

⁴¹(i) Some think of themselves as offering mathematical models for an abstract system, call it ENGLISH, which relates atomic symbols of the language with semantic values, and describes their systematic combination to yield truth-conditions for sentences. (ii) Others take themselves to be studying what must be (at least tacitly) *known* by competent of (say) English— viz. rules that allow one to generate and interpret infinitely many sentences simply by understanding the meanings of their parts. (iii) Still others are concerned with the mechanisms that implement the structure of language in human minds.

⁴² As Paul Elbourne once put it when certain philosophers expressed concern about the metaphysical commitments of semantic theories: “I’m just trying to model what goes on in people’s heads. I’m not sure I understand what you’re doing” (Elbourne, p.c.). See also Chomsky (1975, 1986).

⁴³ ‘We are concerned only to say what system of expectations a normal member of the language-using population must have. We need not engage in psychological speculation about how those expectations are generated’ (Lewis 1975, 25).

⁴⁴ Among those who doubt that semantic theory has straightforward consequences for ontology are Azzouni (2004), Brown (2009), Sider (2011), Williams (2011).

representation... at which arguments at LF are paired with entities of mental representation, this further level then entering into “real semantic interpretation”.

I would put things somewhat differently. We should not say, either in English or in a metaphysically perspicuous language, that flaws and shadows are ‘entities of mental representation’. (In English they are ‘out there in the world’; in Ontologese they are not mentioned at all.) But I will endorse a nearby point: the smoothest theory of meaning might have to quantify over flaws, while the metaphysically perspicuous sentences that *account* for that theory of meaning will not.⁴⁵ Thus, for example, suppose the semantic theory says:

A sentence of the form ‘The F is G’ is true iff there is a unique object that has the property expressed by ‘F’ and it has the property expressed by ‘G’.

This is intended to cover cases where the NP picks out a flaw, and also cases where it picks out a coat. Clearly, the sentence of a reductive theory that accounts for ‘The flaw has mass’ will look very different from the sentence that accounts for ‘The coat has mass’. For this reason, it seems clear that any articulation in such a theory for what it takes *in general* for an English sentence of the form ‘The F is G’ to be true will be highly complex. Moreover, the result would be of no practical use in predicting and explaining people’s linguistic capacities. In parsing the two uses of ‘The F is G’, the language organ is completely insensitive to the vast structural differences in what it takes for them to be true, from a fundamental point of view. (Likewise, to use the language of the instrumentalist, for the abstract rules that best systematize the conventions of language use for noun phrases.) So when it comes to predicting and explaining people’s linguistic capacities, we should ignore these differences.

It’s not until we have to account, in our theory of reality, for talk about flaws and coats—including the talk that goes on in the semantic meta-language—that we find ourselves having to treat facts about flaws very differently from facts about coats. But all that awkwardness enters in where it should be expected: in an account of an objectively complex phenomenon using an

⁴⁵ See the related distinction Ted Sider makes between what he calls ‘metaphysical semantics’ and ‘linguistic semantics’ (Sider 2011, §7.4). See also Brown (2009), Williams (2011) and a related discussion of the ‘ontological commitments’ of a semantic meta-language in Azzouni (2004): 53-62.

objectively simple language.⁴⁶ Thus, for the sake of a smooth semantics, it is arguably best to give the semantic theory in a language that quantifies over things like stocks and bonds, flaws and shadows. But this has no more metaphysical significance than the fact (which reducers already acknowledge) the people frequently say true things while quantifying over flaws and shadows and the like.

As a second example, suppose that the most systematic model of natural language posits implicit quantification over events (or situations) at the level of logical form.⁴⁷ If this is right, the most elegant theory of meaning should assign a domain of events for these event quantifiers. But suppose our metaphysical theory does not have the ontological resources to directly supply a domain whose elements can be taken straightforwardly as values for event quantification. In that case, it will turn out that the most elegant semantic theory is not stated in a very metaphysically perspicuous language.

Finally, consider sentences about fictional objects, such as ‘Santa doesn’t exist’, ‘Santa is jolly’, and ‘There is someone that children think comes down chimneys, namely Santa’. Even though ‘Santa Claus’ appears to be behaving just like an ordinary proper name, attempts to account for such sentences without appealing to fictional entities require treating that expression as behaving very differently in the three sentences. And it is common to compare various metaphysical theories of fictional objects in part based on how uniformly it can treat these sentences.⁴⁸ However, on the division of labor I am suggesting, this conflates two distinct projects. The fact that all these Santa-sentences compose with perfect smoothness can arguably be explained most straightforwardly by the hypothesis that our language organ is tacitly somewhat

⁴⁶ And this is not to say that there is no explanation for why we use sentences of the same structure to represent objectively very different situations. But it will be part of a deeper theory of cognition than anything that compositional semantics aims for. Moreover, arguably this explanatory cost is less than that paid by the anti-reductionist elsewhere in her theory— not only does she need tornados, but she must posit a brute metaphysical *connection* between tornados and the underlying air particles. If this is right, the explanatory bump in the carpet is just being shifted—from an account of why we *represent* whirling air particles as unified entities, to an account of what *makes* them unified entities.

⁴⁷ Thus an utterance of ‘Fido barks’, along with its attendant unvoiced structure, can be represented by ‘ $\exists e[\text{barks}(e) \ \& \ \text{agent}(e) = \text{Fido}]$ ’. Treating a verb as a predicate of events, linked to its arguments by thematic roles, will yield a semantic tree for ‘Fido barks’ with no less than four terminal nodes and three compositional operations. See e.g. Schein (2002, 2012).

⁴⁸ See, for example, Parsons 1980, Thomasson 1999.

Meinongian.⁴⁹ If so, a compositional account of meaning whose structure matches that of the language organ would proceed using a Meinongian language. And again, it is not until we must account for talk about fictional individuals in our theory of reality that we find ourselves having to treat the various ‘Santa’-sentences in fundamentally different ways.

We have said that a theory of reality should be simple where the world is simple, and complex where the world is complex. And I know of no reason to doubt that the workings of our language organ, or of our linguistic practices, are much more objectively complex than would be evinced by the structure of any useful theory of meaning. If they are, then trying to articulating a systematic semantics in the language of metaphysics may be very much like trying to articulate a workable theory of genetic inheritance in the language of fundamental physics. (This is not to deny that, *ceteris paribus*, we should prefer theories on which a given phenomenon turns out to be objectively simpler. But we assumed at the outset that T3 and T4 are equally simple theories *overall*. Our question is whether English-friendliness—admitting of a relatively systematic procedure for identifying reductions of English sentences—is a virtue distinct from *overall* simplicity of theory.)

This conclusion ties in with an earlier point about explanations in the special sciences. The goal of a special science is uncover regularities and causal connections allowing us to navigate phenomena that we may be interested in for various pragmatic reasons, even if they are objectively highly complex. And the resulting theories must be sufficiently tractable that they can be employed in scientific settings to predict and explain the relevant phenomena. But this in turn may only be possible because the theory is expressed in a language whose expressions are relatively gerrymandered.⁵⁰

In short: NL-friendliness is not a theoretical virtue. There is, in the end, just one criterion of simplicity for theories of reality: prefer those theories on which the world as a whole is simpler, even if they count natural language meaning as among the world’s highly complex phenomena.

⁴⁹ Contrast the semantic acrobatics undertaken by theories that identify Santa Claus with abstract objects—‘Santa Claus is fat’ has one kind of semantic interpretation, while ‘Bob is fat’ has another. Meanwhile ‘pretense’ theories face the problem (among others) that no pretense seems to be going on at all when we say ‘Santa Claus doesn’t exist’. See Ludlow 1999 pg. 71, and also Brown, 2009, ch. 2.

⁵⁰ (Even so, such theories may actually be providing the most objectively simple explanations possible, given the objective complexity of their explananda. In this sense the facts they offer as explanations—whether articulated concisely in the language of the special science or laboriously in the fundamental language— may be the most explanatory facts available.)

9. Recap

We've surveyed a variety of competing criteria for assessing how well a reductive theory saves the appearances.

We began with the requirement that a theory can account for all the facts. I argued this requirement applies only to the kind of perfect theory that serves as a regulative ideal for flesh-and-blood metaphysicians. Theories of the sort we actually present to each other are indeed failures if not *completable*— i.e. capable in principle of being fleshed out so as to account for all the facts. But the dialectical situation in metaphysics hardly requires proponents of a theory to shoulder the burden of *establishing* that it's completable.

We also considered some potential ideas about what it takes to 'account' for all the facts in the relevant sense. How we answer this question depends on our view about the aim of a metaphysical theory: whether we are (in Fine's terms) E-theorists or D-theorists. Is the aim of a theory to say, for every fact, *what it is* for that fact to be the case? Or is it sufficient to say *what makes it* the case? Either way, we face another set of difficult questions when we ask which putative facts are actually accounted for by a given theory. This is where interpretive charity comes in—a desideratum not for choosing between theories of reality, but for asking which natural language sentences come out true if we take a given theory to be the correct metaphysics. I argued that interpretive charity is not a *sui generis* desideratum, but amounts an application of the principle of theoretical simplicity.

Finally, we examined a few varieties of theoretical simplicity that seem relevant to the project of saving the appearances: ideological simplicity, ontological simplicity, ordinary-fact simplicity, and NL-friendliness. I argued that the latter two are not theoretical virtues at all, unless there is a special reason to think that the sorts of things we concern ourselves with as human beings are among the simpler aspects of reality. Meanwhile, the first two criteria of simplicity boil down to this: first, favor theories according to which the world is a simpler place overall, and then favor those whose linguistic structure matches the structure of the reality they describe.

Works Cited

- Armstrong, David 2004. *Truth and truth-makers*, Cambridge: Cambridge University Press.
- Azzouni, Jody. 2004. *Deflating existential consequence: a case for nominalism*. Oxford: Oxford University Press.
- Bigelow, John. 2009, "Truthmakers and Truthbearers," in Robin Le Poidevin, Peter Simons, Andrew McGonigal and Ross P. Cameron (eds.), *The Routledge Companion to Metaphysics*, London: Routledge, 389–400.
- Brown, Joshua. 2009. 'Deflating referential semantics' in *From semantics to metaphysics*. Unpublished dissertation.
- Cameron, Ross. 2008 "How to be a truthmaker maximalist," *Noûs*, 42: 410–21.
- Chalmers, David. 2011. 'Verbal disputes'. *Philosophical Review*, 120(4): 515-566.
- Chomsky, Noam, 1986. *Knowledge of language: its nature, origin, and use*. New York: Praeger.
- . 1981. *Lectures on government and binding: The Pisa lectures*. Foris Publications: Dordrecht.
- . 1975. *Reflections on language*. New York: Pantheon.
- Dorr, Cian. 2005. 'What we disagree about when we disagree about ontology.' In *Fictionalism in metaphysics*, ed. M. Kalderon. Oxford: Oxford University Press, 234 – 286.
- Fine, Kit. 2013. 'Fundamental truth and fundamental terms'. *Philosophy and Phenomenological Research* 87(3): 725-732.
- . 2012. 'Guide to ground'. In Fabrice Correia & Benjamin Schnieder (eds.), *Metaphysical Grounding*. Cambridge University Press. 37-80.
- . 2001. 'The question of realism'. *Philosopher's Imprint*, 1.
- Fodor, Jerry A. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
- Hirsch, E. 1982. *The Concept of Identity*, Oxford: Oxford University Press.
- Lewis, David. 1986. *On the plurality of worlds*. Oxford: Blackwell.
- . 1983. "New work for a theory of universals." *Australasian Journal of Philosophy* 61: 343–77

- . 1975. ‘Languages and Language’. in Keith Gunderson, ed, *Minnesota Studies in the Philosophy of Science*, 7. Minneapolis: University of Minnesota Press.
- . 1974. ‘Radical interpretation.’ *Synthese*, 23, 331-344.
- Ludlow, Peter. 1999. *Semantics, tense, and time: An essay in the metaphysics of natural language*. Cambridge, MA: MIT Press.
- Melia, Joseph. 1995. ‘On what there’s not’. *Analysis*, 55, 223–9.
- . 2000. ‘Weaseling away the indispensability argument’. *Mind*, 109(435), 455-480.
- . 2005. ‘Truthmaking without truthmakers’ in Beebe, Helen and Julian Dodd, eds. *Truthmaking: the contemporary debate*. Oxford: Oxford University Press.
- Mellor D. H. 2009. ‘Truthmakers for what?’ in H. Dyke, ed. *From Truth to Reality: New Essays in Logic & Metaphysics*. London: Francis & Taylor.
- . 2003, “Real Metaphysics: Replies,” in Lillehammer & Rodriguez-Pereyra (eds.) *Real Metaphysics: essays in honour of D.H. Mellor*, London: Routledge.
- Merricks, Trenton. 2007. *Truth and ontology*, Oxford: Clarendon Press.
- Parsons, Terence. 1980. *Nonexistent objects*. New Haven: Yale University Press.
- Price, Huw. 2009. Metaphysics after Carnap: the ghost who walks? In *Metametaphysics*, eds. D. Chalmers, D. Manley and R. Wasserman. Oxford: Oxford University Press
- Quine, W. V. 1951. Two dogmas of empiricism. *The Philosophical Review* 60:20-43.
- Rayo, Agustin. 2008. ‘On specifying truth-conditions’. *Philosophical Review*, 117, 385–443.
- Schaffer, Jonathan. 2010. ‘On what grounds what’. In *Metametaphysics*, eds. D. Chalmers, D. Manley and R. Wasserman. Oxford: Oxford University Press.
- Schein, Barry 2002. ‘Events and the semantic content of thematic relations.’ In Gerhard Preyer & Georg Peter, eds., *Logical Form and Language*. Oxford: Oxford University Press.
- . 2012. ‘Event semantics’. Forthcoming in Delia Graff Fara & Gillian Russell, eds., *The Routledge Companion to Philosophy of Language*. Oxford: Routledge.

- Sellars, Wilfrid 1962. 'Philosophy and the scientific image of man'. in *Frontiers of Science and Philosophy*, ed. Robert Colodny/ Pittsburgh: University of Pittsburgh Press. Pgs 35-78.
- Sider, Theodore. 2011. *Writing the book of the world*. Oxford: Oxford University Press.
- . 2009. 'Ontological realism'. In *Metametaphysics*, eds. D. Chalmers, D. Manley and R. Wasserman. Oxford: Oxford University Press.
- Simons, P. 2008. "Why the negations of false atomic sentences are true," in T. De Mey & M. Keinänen (eds.), *Essays on Armstrong*, Acta Philosophica Fennica, 84: 15–36.
- Soames, Scott. 2009. Ontology, Analyticity, and Meaning: the Quine-Carnap Dispute. In *Metametaphysics*, eds. D. Chalmers, D. Manley and R. Wasserman. Oxford: Oxford University Press
- Sosa, Ernest. 1993. 'Putnam's pragmatic realism.' *Journal of Philosophy* 60 (12): 605-626.
- Thomasson, Amie 1999. *Fiction and metaphysics*, Cambridge: Cambridge University Press.
- Williams, J. Robert G. 2010. 'Fundamental and derivative truths'. *Mind*, 119(473), 103–141.
- . 2011. 'Requirements on reality'. unpublished ms. available at www.personal.leeds.ac.uk/~phljrgw/wip/reqrealityrevised2.pdf
- Wilson, Jessica. 'No work for a theory of grounding'. unpublished ms. available at <http://individual.utoronto.ca/jmwilson/>