

## Is Kant's Critique of Metaphysics Obsolete?

Updated 11/1/23

*Draft—do not cite or circulate without permission*

### Abstract

I raise a problem about the possibility of metaphysics originally due to Kant: what explains the fact that the terms in our metaphysical theories (e.g., ‘property’, ‘grounding’) refer to entities and structures (e.g., properties, grounding) in the world? I distinguish a meta-metaphysical view that can easily answer such questions (‘deflationism’) from a meta-metaphysical view for which this explanatory task is more difficult (which I call the ‘substantive’ view of metaphysics). I then canvass responses that the substantive metaphysician can give to this Kantian demand for an explanation of reference in metaphysics. I argue that these responses are either inadequate or depend, implicitly or explicitly, on the idea of ‘joint carving’: carving at the joints is part of the explanation of reference-facts quite generally and our metaphysical terms in particular refer because they carve at the joints. I examine Ted Sider’s recent work on joint carving and structure and argue that it cannot fill the explanatory gap. I conclude that this is reason *ceteris paribus* to reject the substantive view of metaphysics. Kant’s critique, far from being obsolete, applies to the most cutting-edge of contemporary meta-metaphysical views.

### §1. Is Kant’s Critique of Metaphysics Obsolete?

In 1790 Kant broke his policy of not replying in print to criticisms of his work—a policy made necessary by his awareness of his advanced age and the significant work that still remained to be done on the Critical system—by publishing an essay titled *On a discovery according to which all critique of reason has been rendered obsolete by a previous one*.<sup>1</sup> This essay is a response to Johann August Eberhard, who had claimed that Kant’s investigation into the limits of reason had already been carried through

---

<sup>1</sup> *Über eine Entdeckung nach der alle neue Kritik der reinen Vernunft durch eine ältere entbehrlich gemacht werden soll.* ‘Entbehrlich’ is better rendered as ‘superfluous’ (as is done in Henry Allison’s translation in Kant 2002) but I have chosen ‘obsolete’ because it better fits the theme of this essay.

by Leibniz and Wolff.<sup>2</sup> Kant's critique had allegedly been rendered obsolete, according to his somewhat sarcastic title, before it was even written. In this paper I want to consider whether a part of Kant's critique of reason, in particular, his critique of metaphysics, has been rendered obsolete, not by its predecessors, but by its successors, in analytic metaphysics.

As every student of philosophy knows, Kant took himself to have shown in the *Critique of Pure Reason* that traditional metaphysics is beyond our cognitive limits.<sup>3</sup> Nonetheless, metaphysics has enjoyed a splendid resurgence in analytic philosophy over the last forty years. Hundreds of books and articles have been written on subjects that would have been familiar to Kant and his contemporaries: the nature of possibility and necessity, causation, grounding, particulars and their properties. Even the principle of sufficient reason is back on the table.<sup>4</sup> This has given rise to a feeling among some philosophers that analytic metaphysicians are, mistakenly, returning to a kind of metaphysics that Kant showed to be impossible.<sup>5</sup> The converse reaction, felt by those more sympathetic to the return of metaphysics, is that this underscores the irrelevance of Kant to contemporary metaphysics: metaphysics is not impossible (witness the latest work), so Kant's critique must not have been as 'all-destroying' as some have taken it to be.<sup>6</sup>

There are several good reasons to think that Kant's critique of metaphysics simply does not apply to the resurgence of metaphysics in analytic philosophy in the past decades. These reasons unite around a common theme: what Kant meant by metaphysics is more ambitious in its scope and more arrogant in its claims to knowledge than contemporary metaphysics. If Kant showed anything, it is that a particular brand of metaphysics (metaphysics in the continental rationalist tradition, say) is

---

<sup>2</sup> Eberhard's original essay and Kant's reply are reprinted in Kant 2004.

<sup>3</sup> Matters are a bit more complicated. The metaphysics that Kant takes to be beyond our cognitive limits is metaphysics of 'things in themselves,' things as they are, independently of whether we can ever experience or know them. But Kant retains a place for something else he calls 'metaphysics,' synthetic a priori cognition of how things appear to us in experience, for which the *CPR* is merely the preparatory critique (Bxxvi), not the completed system (Bxlili). Since the metaphysics that has flourished in analytic philosophy since David Lewis is, in Kantian terms, about things in themselves, Kant's arguments against such a metaphysics potentially apply.

<sup>4</sup> See Pruss 2009 & Della Rocca 2010.

<sup>5</sup> E.g., van Fraassen 2002, 2–4.

<sup>6</sup> See Williamson 2007, 19; Lowe 1998, 1–8.

untenable, or that metaphysics cannot obtain a certain kind of epistemic status (a priori knowledge, certainty, the ‘secure path of a science,’ etc.).<sup>7</sup> Some such reasons are as follows:

1. *Analytic/Synthetic*. As Kant presents his critique of metaphysics in the Introductions to both editions of the *CPR*, the central issue is the possibility of our having synthetic a priori knowledge. However, ever since Quine 1951 many philosophers have been suspicious that there is such a thing as the analytic/synthetic distinction and have thus been skeptical of any philosophical project that depends on it.<sup>8</sup>
2. *A Priori*. The other half of Kant’s critique of the possibility of synthetic a priori knowledge, of course, is that the knowledge in question must be a priori. Kant assumes that metaphysics, if there is such a thing, must be completely independent of experience: metaphysics ‘elevates itself entirely above all instruction from experience’ (Bxiv).<sup>9</sup> But contemporary metaphysicians do not conceive of their project in such starkly a priori terms. Many take themselves to be generalizing or systematizing ordinary judgments about ordinary empirical objects. Some explicitly rely on our best current physics, which is partly based on empirical evidence and observation.<sup>10</sup> So if Kant’s arguments about metaphysics rest on the assumption that metaphysics must be strictly a priori, they seem outdated.<sup>11</sup>

---

<sup>7</sup> Some of these points are raised by Lowe 1998.

<sup>8</sup> The fate of analyticity does not end there. The analytic-synthetic distinction still has its friends (Russell 2008) and its foes (Williamson 2007, 48–133). One’s reaction to the role of the analytic/synthetic distinction in Kant’s project likely depends on one’s meta-metaphysical commitments. Adherents of ‘substantive’ meta-metaphysics (e.g., Lewis, Sider, Schaffer; see below) would agree with Kant that metaphysics is ‘synthetic,’ but think his critique fails because it depends on the assumption that metaphysics must be synthetic and a priori (point 2 in the main text). Adherents of ‘deflationary’ meta-metaphysics (e.g., Thomasson; see below) would object against Kant that metaphysics is largely analytic. So regardless of where one stands in the contemporary landscape, Kant’s critique of synthetic a priori metaphysics is liable to seem out of date.

<sup>9</sup> The *Critique of Pure Reason* is cited in the customary format: the page in the 1781 edition (A), followed by the page in the 1787 edition (B). The *Critique* is quoted from Kant 1998.

<sup>10</sup> E.g., Maudlin 2007.

<sup>11</sup> Sider 2011, Williamson 2007, Ladyman and Ross 2007.

3. *Big 'M' metaphysics.* This is a slightly more nebulous issue, but the basic idea is that metaphysics, as Kant conceives it, is a grand and ambitious affair: 'Metaphysics has as the proper end of its investigation only three ideas: God, freedom, and immortality' (B395n).<sup>12</sup> Contemporary analytic metaphysics is a comparatively modest business. While there is extensive work on the metaphysics of free will, and some work on the metaphysics of theism and personal immortality, few (if any) contemporary metaphysicians would claim these are the only 'proper ends' of their investigation.
4. *'Epistemology first.'* Kant's critique of metaphysics appears to assume an 'epistemology first' methodology: first we determine the limits of our epistemic capacities, then we determine whether metaphysics lies within that scope. From the 1780s to today this can seem to beg the question against the metaphysician by assuming controversial epistemic claims the metaphysician would reject, such as the intuition-dependence of our knowledge (see 7 below). Contemporary metaphysics, and philosophy more generally, tends to pursue a more 'holistic' approach to philosophical theorizing and justification. We start with some plausible assumptions in metaphysics, epistemology, philosophy of mind, etc. and through a gradual process of sifting objections, replies, etc. we find 'reflective equilibrium,' the overall most rationally satisfactory combined set of theses in epistemology, metaphysics, etc. Epistemology has no privilege or priority over metaphysics; they are both subject to the 'dialectical free-for-all.'<sup>13</sup>
5. *Apodictic certainty.* Related to the issue of the a priori, Kant conceives of metaphysics as aiming at apodictic certainty. For these reasons, he claims that metaphysics cannot rest on hypotheses or judgments about what is probably the case.<sup>14</sup> But few contemporary

---

<sup>12</sup> 'There was a time when metaphysics was called the queen of all the sciences, and if the will be taken for the deed, it deserved this title of honor, on account of the preeminent importance of its object' (Aviii); 'metaphysics is also the culmination of all culture of human reason' (A850–1/B878–9).

<sup>13</sup> See Lewis 1986, 108–115; Sider 2011.

<sup>14</sup> Axv, A774/B802.

metaphysicians would claim to have demonstrated their conclusions with certainty; the most that many metaphysicians would claim is that theirs is the theory overall best supported by various considerations (e.g., elegance, adequacy to the data, etc.).

6. *Supersensible.* Metaphysics, Kant writes, ‘is the science of progressing by reason from knowledge of the sensible to that of the super-sensible.’<sup>15</sup> The ‘supersensible’ does not refer to the realm of abstracta but to a realm of concrete beings that cannot be sensed by us (so-called ‘noumena’). As a result, Kant does not think that natural science, even at its ideal limit, could constitute metaphysics, because natural science, even micro- and astrophysics, concern themselves with the spatiotemporal world. On this point, Kant’s conception of metaphysics may seem especially irrelevant to contemporary practice. First of all, the idea that metaphysics has to do with the supersensible is virtually absent from the contemporary scene. Indeed, many would deny that supersensibles (in Kant’s sense) even exist in the first place.<sup>16</sup> Secondly, many contemporary metaphysicians take the project of metaphysics to be continuous with empirical natural science. If Kant’s point is that a priori knowledge of the supersensible is impossible, the contemporary metaphysician might agree with him, but feel that that this has little or nothing to do with metaphysics as she and others practice it.
7. *Intuition-dependence.* Finally, it might be thought that Kant’s critique of metaphysics depends both on his own technical notion of ‘intuition’ (*Anschauung*) and his views about the dependence of our knowledge on intuition. Since the first is tied up with Kant’s arguments in the *Transcendental Aesthetic*, which, notoriously, appears invalid,<sup>17</sup> and the

---

<sup>15</sup> Kant 2002, 353.

<sup>16</sup> The closest contemporary analogue would be the view that many of the paradigm objects of metaphysics, e.g., properties, are not concrete but abstract, and hence are neither spatiotemporally located nor causally efficacious. However, contemporary metaphysicians do not conceive of abstracta, such as properties, as ‘supersensible’ beings that ground or lie beneath or behind perceptible objects.

<sup>17</sup> The apparent invalidity is known as ‘Trendelenburg’s gap.’ For recent attempts to fill it see Hogan 2009 and Rosefeldt 2013; Hogan’s paper also contains extensive references to earlier work on the topic.

second is highly controversial in its own right, this might not seem like a promising basis for a critique of contemporary metaphysics. Things look even worse for Kant when we remember that his doctrine of intuition is closely tied to his philosophy of mathematics, according to which mathematics is not about mind-independent reality but about the spatiotemporal form of our intuition, a view widely rejected by philosophers of mathematics.<sup>18</sup> *If Kant's critique of metaphysics depends on this stuff*, the contemporary metaphysician might reply, *then surely I am justified in ignoring it*.

To summarize, then, contemporary metaphysics is less a priori, less dependent on the analytic-synthetic distinction, less committed to an 'epistemology first' methodology, less ambitious in the scope of its theorizing, less ambitious in the epistemic status it claims for those theories, and more continuous with the empirical sciences. 'Metaphysics,' as conceived by Kant, is vastly more ambitious than what now goes under that name in analytic philosophy; in rejecting that more ambitious project, contemporary philosophy might even be thought to have *assimilated* the insights of Kant's critique of metaphysics, such as they are.

Nonetheless, I think that Kant's critique of metaphysics has much left to teach us, that contemporary metaphysics has not absorbed its insights, and that when these insights are separated from some of Kant's other philosophical commitments, they are shown to pose a powerful challenge to metaphysics, even in its more modest contemporary forms. This paper is part of a much larger project<sup>19</sup>; here I can only present one aspect of Kant's critique and my argument that contemporary metaphysicians have failed to answer it. This is not a work of Kant exegesis, but an attempt to apply some Kantian ideas to contemporary metaphysics. However, in articulating the Kantian challenge to metaphysics I will sometimes express myself in terms of what Kant says, argues, etc. These claims

---

<sup>18</sup> Indeed, one can think of the development of philosophy of logic, mathematics, and semantics in the 19<sup>th</sup> and early 20<sup>th</sup> century as attempts to do without the mysterious Kantian notion of a 'form of intuition' (Coffa 1991). Furthermore, Kant's theory of forms of intuition is, arguably, simply disproven by later technical developments (Friedman 1992, 55–95); for a dissenting view see Hanna 2006, 287–340.

<sup>19</sup> For other parts of the project see Stang Forthcoming-a, Forthcoming-b, and 2023.

could be replaced, without significant loss of meaning, with claims about what ‘the Kantian’ says, argues, etc.<sup>20</sup> It is not crucial to the argument of this paper that I am correctly interpreting what Kant himself actually wrote or thought.<sup>21</sup>

In Section 2 I explain the Kantian objection I will be examining in this paper: the metaphysician is committed to certain terms in her theory referring to entities and structures in the world but cannot explain how this is possible. I will refer to this as the ‘explanatory gap’ objection. In Section 3 I refine my target by pointing out that ‘deflationary’ views in meta-metaphysics can easily explain the reference of metaphysical terms. The explanatory gap objection thus does not apply to those views, but to what I call ‘substantive views’ in metaphysics. Because there are deflationary views that can fill the explanatory gap, if the substantive metaphysician cannot do so, this constitutes a real problem for such a view. This paper is *not* an argument for deflationism; to constitute such an argument, it would need to assume something which I at any rate reject, that deflationism is the *only* meta-metaphysical view that can fill the explanatory gap.<sup>22</sup> Several substantivist strategies for dismissing the demand for an explanation of metaphysical term reference are considered, and rejected, in Section 4. In Section 5 I canvass contemporary metaphysics for answers to this Kantian objection and argue that none is forthcoming. The closest thing to a response is the notion of ‘structure’ in Sider 2011. Section 6 is devoted to exploring structure and ‘joint carving’ and how they might be marshaled to answer the Kantian explanatory gap objection. I argue that Sider’s explanation fails; the explanatory gap re-emerges on his view as well. I conclude that this is a reason *either* to reject the substantive view of metaphysics altogether *or* to supplement it with enough additional metaphysics to explain metaphysical term reference. Further pursuing either option lies, however, outside the scope of this essay. Section 7 responds to two objections, and section 8 concludes.

---

<sup>20</sup> In the same way that one might talk about what ‘the Humean’ says about causation.

<sup>21</sup> Though it is backed-up by a textually and historically informed interpretation; see my papers cited at the end.

<sup>22</sup> This is part of an in-progress monograph, *How is Metaphysics Possible?*, in which I argue that neither inflationism or deflationism can fill the explanatory gap and explore whether Kant’s own transcendental idealism can do better.

Before continuing I want to make clear that it is not a commitment of Kant's critique of metaphysics, nor of this paper, that metaphysics is nonsense or can be dispensed with. The aim of Kant's critique of metaphysics is to explain how metaphysics is possible and what its appropriate limits are. The aim of this paper is to argue that a certain (substantive) conception of metaphysics is problematic, and thus to motivate the search for an alternative conception of metaphysics. Only the former project is explicitly undertaken here, however.

## §2. The Reference of Metaphysical Concepts

Allow me one Kant quotation:

I noticed that I still lacked something essential, something which, in my long metaphysical studies, I, as well as others, had failed to consider, and which in fact constitutes the key to the whole secret of metaphysics, hitherto hidden even from itself. I asked myself this question: what is the ground of the relation [*Beziehung*] of that in us which we call 'representation' to the object?<sup>23</sup>

This is from Kant's 1772 letter to Marcus Herz, in which he announces the project that would be published, nine years later, as the *Critique of Pure Reason*. It is not immediately clear from context what 'relation' between representations and objects Kant has in mind but I think it should be understood as *reference*.<sup>24</sup> Kant's question then becomes: what explains ('on what ground rests') the reference of our mental representations to objects?<sup>25</sup> In the rest of the letter, Kant goes on to say that he finds this relation (which we are taking to be *reference*) relatively unproblematic in the case of mathematical concepts and empirical concepts (e.g., natural kind concepts like <*gold*>). He cannot find a satisfactory explanation of the fact that the concepts of metaphysics (e.g., <*substance*>, <*causation*>, <*reality*>, etc.) refer to their objects. This is the problem that becomes the problem of the 'objective validity' of the categories, the problem that is supposed to be solved by the Transcendental Deduction and the Schematism. So the

---

<sup>23</sup> Kant 1999, 133. Translation slightly modified by author.

<sup>24</sup> I won't defend the interpretive point here, but see Stang 2018.

<sup>25</sup> I do not want to get into the details of Kant exegesis here, but I do think in the specific context of contemporary metaphysics it is not wrong to read '*Beziehung*' here as reference.



question Kant first formulated in 1772 in this letter to Herz ('the key to the whole secret of metaphysics, hitherto hidden even from itself') becomes one of the central questions of the whole *CPR*: why do metaphysical concepts refer?<sup>26</sup>

This, however, is a question we could just as well raise for contemporary metaphysics. Contemporary metaphysicians might be more comfortable talking about the reference of words or terms in metaphysical theories (perhaps understood as sets of sentences). While I do think that Kant was right to raise the problem at the level of thought (concepts) not at the level of language (words), for the purposes of this essay I will formulate the question linguistically: if any metaphysical theory has a truth-value, given a referential semantics, the referential terms in that theory refer to entities and structures in the world.<sup>27</sup> What explains that fact?

By a 'referential semantics' I mean one that contains familiar Tarski-style recursive truth-conditions for atomic sentences, quantifiers, logical connectives, etc. We have a domain  $D$  and a function  $R$  (the *reference* function) that maps certain expressions in the language to entities and structures in the domain that contribute to the truth-values of sentences in which those expressions appear. In particular,  $R$  maps each constant in the language to an object in  $D$ , and each predicate in the language to an extension, a set of ordered pairs of objects in  $D$ . An atomic sentence  $F(a_1 \dots a_n)$  is true just in case  $\langle R(a_1) \dots R(a_n) \rangle \in R(F)$ . A quantified sentence  $\forall x F(x)$  is true just in case all of the objects in  $D$  are in  $R(F)$ . We can then define truth for molecular sentences using the usual clauses for truth-connectives.

---

<sup>26</sup> Kant's famous question 'how is the science of metaphysics possible?' can be separated into several different questions, depending on what 'possibility' is at stake: (1) semantic (how is it possible for metaphysical concepts to refer?); (2) epistemic (how is it possible for us to acquire knowledge in metaphysics?); (3) logical (how is metaphysics logically consistent, given its alleged commitment to Antinomial conflicts?); and (4) scientific (how is it possible for metaphysics to be a *Wissenschaft*?). While much of Kant scholarship has focused on (2), this essay concerns (1). Stang Forthcoming-b concerns (4). See Kriegel (2013) for a contemporary version of the Kantian challenge to the epistemic possibility of (revisionary) metaphysics.

<sup>27</sup> I use the expression 'entities and structures' to accommodate Ted Sider's view that structure is not itself an entity, not something to be 'quantified over' (Sider 2011, 100–123).

While R itself only takes constants and predicates and arguments, we can naturally extend this notion of reference beyond items of these syntactic categories. If D is the domain of quantification then D makes a semantic contribution to the truth-conditions of quantified sentences. In an extended sense, quantified sentences ‘refer’ to D; D is, so to speak, the ‘referent’ of the universal quantifier  $\forall$ . We can ask, for instance, why  $\forall$  has this meaning, rather than another.<sup>28</sup> We can ask why  $\forall$  refers, in the extended sense, to D, rather than some other domain (perhaps a subset of D, if D is a set).<sup>29</sup>

In this way, questions of reference can also be applied to terms in metaphysical theories (henceforth, metaphysical terms) that are not naturally thought of as referring to objects, such as quantifiers and modal terms. If modal operators are treated as quantifiers over possible worlds, then the discussion of quantification above applies and we can ask how our modal operators succeed in referring to the domain of possible worlds. However, some metaphysicians treat modal operators as primitive features of reality. Nonetheless, our toy model of reference can be generalized to account for such ‘modal primitivists,’ as long the modal primitivist includes in her semantic theory some clause of the form:

- (1) ‘Possibly  $p$ ’ is true if and only if it is possible that  $p$ .<sup>30</sup>

So if modal features of reality (the fact that it is possible that  $p$ ) are to contribute to the truth-conditions of sentences involving modal expressions (‘possibly’), then, in the extended sense, those modal operators ‘refer’ to these modal features. So the very rough picture of metaphysical terms

---

<sup>28</sup> My description of D as the meaning of the quantifier is only intended to set up this question. I am not trying to take sides in the debate about whether the domain, or the rules of inference, or some combination of both, constitutes the meaning of the quantifier.

<sup>29</sup> My description of it as a ‘domain’ notwithstanding, I am not assuming that D is a set, because I do not want to exclude the possibility of absolutely unrestricted quantification, on which the ‘meaning’ of the quantifier is not anything ‘set-sized,’ but absolutely everything there is. See Williamson 2003 for a defense of absolutely unrestricted quantification. Sider 2009 contrasts different ‘meanings’ the quantifiers might have, without assuming these meanings are sets.

<sup>30</sup> This might appear to conflate modal primitivism, a metaphysical view, with a semantic view about the modal operator ‘possibly.’ I do not think it does. For the modal primitivist, if she adopts a broadly referential semantics, takes herself to be speaking about these primitive modal features of reality, and then my question is: what explains this putative semantic fact?

referring can be generalized to metaphysical terms that are neither singular terms (names, definite descriptions, etc.) nor predicates nor quantifiers.<sup>31</sup> This also allows us to extend the notion of ‘reference’ beyond that which is assigned as the semantic value of terms within a model-theoretic semantics. ‘Reference,’ as I will use the term, names the relation between a term in our language, on the one hand, and entities and structures in the world, on the other, in virtue of which those items and structures contribute to the truth-value of sentences containing that term. This is, of course, not intended as a definition of reference, but as an informal explication. I am trying to remain as neutral as possible among different theories of what reference is.<sup>32</sup>

It is up for debate which terms in a theory (or in our language more generally) are referring terms. Names, presumably are. Predicates, quantifiers, modal operators – even logical connectives? That is more controversial. I remain neutral on these matters. My claim is merely that, if the truth of metaphysical theories is a substantive matter (in a sense to be explicated in the next section), then their truth requires referential contact with the world. I do not say which terms in the theory must refer. I assume merely that some core set of terms within the theory refer. Which terms precisely those are is something that will vary, depending upon which metaphysical theory is in question, which semantic analysis of that theory we adopt. Henceforth, ‘metaphysical reference’ will be used as a shorthand for the reference of terms in metaphysical theories.

It is now standard to distinguish *deflationary* views in meta-metaphysics from non-deflationary, or as I will call them, *substantivist* views. I will explain this distinction in more detail in the next section but let me give a rough and preliminary characterization here. According to the deflationist, metaphysical reference can be easily explained by appeal to semantic rules and trivial empirical facts. The substantivist position, as I am using these terms, is simply the denial of this; whatever the explanation of metaphysical reference is, assuming there is one, it is not the deflationary ‘easy’ explanation.

---

<sup>31</sup> Sider 2011, 100–123.

<sup>32</sup> In my defense, neither Williamson 2007, 247–277, nor Sider 2011, 28–41, nor Simchen 2017, 7–11 define what exactly reference is before critically engaging with various views about it.

With this preamble, the Kantian argument can be easily stated:

- (P1) Some metaphysical theory is true.
- (P2) If any metaphysical theory is true then some core set of its terms refer, either to entities in the world or to worldly structure.
- (P3) If substantivism in meta-metaphysics is true then there is no explanation of why the core terms in metaphysical theories refer, either to entities in the world or to worldly structure.<sup>33</sup>
- (C) ∴ Either (i) substantivism is true, but there is no explanation of why metaphysical terms refer, or (iii) substantivism is false.

P1 represents an assumption that all participants in this dialectic, all metaphysicians, must accept; those who think that metaphysics is systematically false, or is not truth-apt, are not the target of this argument.<sup>34</sup> P2 is a conceptual truth about reference, as I have explicated that concept; again, those who think that metaphysics is ‘true’ in some sense that does not require reference, are not the targets of this argument.<sup>35</sup> The argument for P3 will be largely inductive: in Sections 5 and 6 I canvas potential substantivist explanations and argue that none of them succeed. This constitutes solid, but not decisive, evidence that P3 is true. The conclusion presents a dilemma for the substantivist: either embrace unexplained facts of metaphysical reference or reject substantivism. This would not be much of a dilemma if metaphysical reference were inexplicable on any meta-metaphysical view

---

<sup>33</sup> What if the core terms of a metaphysical theory are not distinctively metaphysical (i.e., they are exclusively logical or physical)? I respond to this kind of view below, in Section 4.

<sup>34</sup> Kant is of course familiar with thinkers who reject P1, whom he calls as metaphysical ‘skeptics’ (and among whom he counts Hume), but his argument against metaphysical skepticism is outside the scope of this essay. Closer to our day, one could read Carnap as rejecting P1; metaphysical sentences are meaningless and thus not true (nor are they false). For an account of Kant’s arguments against Hume that connects them to Carnapian empiricism see Westphal 1989, 47–67.

<sup>35</sup> For instance, a metaphysician could adopt an expressivist semantics for metaphysics in which the basic role of metaphysical discourse is to express commitment to, say, a certain conceptual scheme or way of thinking about the world (an historical tradition, etc.). Using strategies made famous by quasi-realists in meta-ethics they could even define a truth-predicate for this very same set of sentences (and even a quasi-realist story about reference). This expressivist option in metaphysics thus constitutes another way to avoid my Kant-inspired argument. Alternately, they could adopt a quasi-realist account of reference, and then they would count as a deflationist, on my way of carving up logical space (see next section.) However, a main strand in contemporary metaphysics adopts a straightforwardly referential semantics about its own discourse, so they at least could not go expressivist. They constitute the target of my argument.

whatsoever. This is why, in the next section, I will argue that the consequent of P3 does not hold for all meta-metaphysical views. Theorists who adopt a deflationary meta-metaphysics can easily explain why metaphysical terms refer. At the end of this essay I will argue that we should prefer a conception of metaphysics on which this unexplained fact does not arise; we should not take the first horn of the dilemma. Whether this should be accomplished by offering a new theory of reference within substantive meta-metaphysics (thus rejecting P3), or embracing a different meta-metaphysics altogether (e.g., deflationism), is an issue I will postpone for future work.<sup>36</sup>

I anticipate that some readers will already want to object that the meta-semantic problem I am raising for metaphysics also applies to the terms in mathematical and physical theories (and, indeed, to the rest of philosophy as well). Thus, my argument might be thought to over-generalize to a problematic extent. I respond to this point in Section 7.

### **§3. Specifying the Target of Kant's Critique: Substantive vs. Deflationary Metaphysics**

Intuitively there is a distinction (though perhaps not a sharp one) between two broad kinds of meta-metaphysical views. Views of the first kind understand metaphysics on the model of physics. There is an objective world 'out there' for us to discover, and our job as metaphysicians is to refer to fundamental aspects of that world and then say true things about them. (A similar speech could be given about physics.) Views of the second kind depict metaphysics as a much more humble affair. Claims in metaphysics are not per se 'about' our conceptual or linguistic scheme (they use, rather than mention, terms like 'property') but they are ultimately downstream of those schemes. Our metaphysical investigations are really investigations of the consequences of the way we conceptualize and speak about reality, not reality *an sich*.

This is obviously a very impressionistic way of making this distinction. There are more precise ways of doing so, but none of them is entirely neutral. For instance, one can distinguish between, on the one hand, views according to which all consistent, expressively complete languages (they have

---

<sup>36</sup> Kant thought that his own 'critical' metaphysics could explain the reference of metaphysical concepts, but that lies outside the scope of this paper.

enough names, predicates, etc. to give a ‘complete’ description) are metaphysically ‘on a par,’ and on the other hand, views according to which some such languages do a better job at ‘carving’ the world ‘at its joints’ (Sider 2011). Alternatively, one can distinguish between views on which ontological questions are ‘easy’ to answer (requiring nothing more than trivial empirical knowledge and knowledge of the semantic rules of one’s language) and views on which ontological questions are ‘metaphysically epistemic,’ that is, require additional, and potentially mysterious, epistemic resources to answer (Thomasson 2014). Each of these ways of drawing the distinction between ‘substantive’ and ‘deflationary’ metaphysics serves certain purposes and captures certain aspects of the intuitive distinction.<sup>37</sup>

My way of dividing up the logical space is slightly different. I distinguish between *deflationary* views, according to which the reference of our metaphysical terms can be easily explained by reflection on the semantic rules of our language, from *substantive* views, according to which metaphysical reference cannot be explained in this way.<sup>38</sup> It is thus the meta-semantic equivalent of Thomasson’s epistemic distinction between (substantive) views on which metaphysics involves ‘epistemically metaphysical’ questions and (deflationary) on views on which it does not. It may seem odd to distinguish meta-metaphysical views on meta-*semantic* grounds, but I think of meta-semantics as a part of metaphysics itself, the part that explains why semantic facts obtain (e.g., why various terms refer).<sup>39</sup> If meta-semantics is part of metaphysics then my meta-semantic objection to

---

<sup>37</sup> But neither is philosophically neutral. Sider’s distinction serves his argument that meta-metaphysical deflationism is a first-order view in metaphysics, the view that the world has no privileged structure, and would be contested by any deflationist who regards the very notion of structure as incoherent or objectionably esoteric (e.g., Hofweber 2009). Thomasson’s distinction motivates her own ‘easy’ approach to ontology (no epistemic mysteries) but would be rejected by a less deflationary theorist who thinks that the eminently empirical method of inference to the best explanation suffices as an epistemology for substantive metaphysics (e.g., Sider himself).

<sup>38</sup> This means that Hofweber 2016’s view about metaphysical terms like ‘property’ ‘number’ and ‘proposition’ counts as deflationary on my lights. Reflection on the semantic rules of our language reveals, according to Hofweber, that these terms do not have a referring function at all. So there is an ‘easy’ (requiring nothing more than reflection on the semantic rules of the language) explanation of why these terms do not refer.

<sup>39</sup> C.f. ‘metasemantics is the business of providing metaphysical explanations of semantic facts’ (Burgess & Sherman 2014, 22).

substantivism is ultimately a metaphysical objection: the substantivist metaphysics lacks, I will argue, the resources to explain the reference of its own metaphysical terms. This also means that the answer to my question, if there is one within substantive meta-metaphysics, will be more metaphysics: a metaphysical explanation of the (semantic) possibility of metaphysics itself.<sup>40, 41</sup>

A wide variety of different meta-metaphysical and meta-ontological views fall under the banner of ‘deflationism’ and I cannot hope to survey all of them here.<sup>42</sup> Instead, I will argue that one prominent version of meta-ontological deflationism, Amie Thomasson’s ‘easy ontology,’ can easily solve this problem. I hope it will be easy to see how to generalize this strategy to other deflationary theories.<sup>43</sup>

Although Thomasson does not, to my knowledge, explicitly address the question of why metaphysical terms refer,<sup>44</sup> it is not hard to construct an answer within her theory. Thomasson’s primary concern is to show that ontological questions (e.g., are there tables, or are there merely particles arranged table-wise?) can be easily answered by a combination of semantic rules that govern the sortal terms in question (e.g., ‘table’) and empirical facts. For instance, the term ‘table’ is governed by this semantic rule, in which it is used rather than mentioned:

(So) If there are particles arranged table-wise then there is a table.

---

<sup>40</sup> This means that my argument had better not be merely a version of the ‘more theory’ response to putative solutions to Putnam’s model-theoretic argument (see Lewis 1984, 224–6). See Section 7 for an explanation of why this is not the case.

<sup>41</sup> This also means that a theorist who combines whatever 1<sup>st</sup> order metaphysical theory one likes (even Sider’s) with a deflationary meta-semantics is not a ‘substantivist’ in my sense, and thus is not the target of my argument here. But that is appropriate. I am offering a meta-semantic objection to a certain view about the meta-semantics of metaphysics.

<sup>42</sup> The locus classicus is Carnap 1950. The Carnapian tradition is carried on by Hirsch 2009 and 2010; Thomasson 2009, 444–471, and 2014; Yablo 2009; Hofweber 2009 and 2016; and the essays in Blatti & Lapointe 2016. I suspect that on my way of classifying positions, Eklund 2009 counts as a deflationist (reference to entities in a pluralist meta-ontology should be ‘easy’), but I am not certain.

<sup>43</sup> The aim of this paper is not to defend Thomasson’s deflationary approach to metaphysics, much less to equate Kant’s highly non-deflationary approach with that of Thomasson. I discuss Thomasson here for two reasons: to show the reader that the problem of explaining the reference of metaphysical concepts is not a problem faced equally by *all* meta-metaphysical views, and to show, by contrast, what explanations are not available to the substantivist. I am thus not endorsing Thomasson’s deflationary view, and so objections to that view do not constitute objections to my argument.

<sup>44</sup> She is more concerned with ontological than with referential questions, though reference is a key component of her solution to ontological problems. See Thomasson 2014, chs. 2-3.

This, in combination, with the fact (agreed upon by all parties to the ontological dispute, let us assume) that there are particles arranged table-wise, entails that there are tables (that tables exist).<sup>45</sup> Thus can we dispense with ‘hard’ ontological debates about whether we should quantify over tables in our best overall theory of the world, whether we would talk about tables in ‘ontologese,’ etc.

This strategy for ‘easily’ answering ontological debates can be just as easily transformed into a strategy for answering questions about reference. Semantic rule  $S_O$  is the object-language correlate (i.e., it *uses* ‘table’) of the following meta-language rule (i.e., it *mentions* ‘table’):

( $S_M$ ) If there are particles arranged table-wise then ‘table’ refers.

The explanation of why ‘table’ refers works along now familiar lines:  $S_M$  and the undisputed empirical facts entail that ‘table’ refers. But notice that the existence of tables was neither part of the explanation of why ‘table’ refers (the right-hand side of  $S_M$ ) nor of why ‘there are tables’ is true (the truth of the right-hand side of  $S_O$ ). That ‘table’ refers is a logical consequence of analytic sentence ( $S_M$ ) and an undisputed empirical truth that does not implicitly or explicitly quantify over tables (according to Thomasson).<sup>46</sup> On Thomasson’s view the existence of tables is playing no work in explaining or grounding the truth of the right-hand side of  $S_M$  (‘table’ refers) because the existence of tables (the right-hand side of  $S_O$ ) is simply the right-hand side of  $S_M$  transposed into the object language. Because she adopts a deflationary view of the quantifiers, semantic rules like  $S_O/S_M$  do not need to be underwritten or explained by some domain of objects to which we could refer without these rules; we can use semantic rules like this to introduce new referring terms into our language without worrying how they relate to our ‘prior’ ontology (the ontological commitments of our language prior to the introduction of these new terms). Likewise, the existence of tables is playing no

---

<sup>45</sup> Thomasson follows Kant and Frege in identifying *what exists* with *what there is* (Thomasson 2014, 85–86).

<sup>46</sup> Understanding the sentence ‘there are tables arranged table-wise’ might be thought to implicitly depend on understanding the concept of a table. If so, we can change the example: ( $S_O$ ) if there are some Fs then there is a mereological fusion of the Fs. The argument can be recast using ‘mereological fusion’ rather than ‘table.’



work in explaining or grounding the truth of the right-hand side of  $S_O$  because, on the disquotational view of truth Thomasson endorses, ‘there are tables’ and ‘*there are tables* is true’ are not separate facts.

I have explained Thomasson’s strategy in the case of ordinary sortal terms like ‘table’ but this strategy can be easily extended to explain the reference of more metaphysically-loaded terms like ‘property,’ provided that these terms are governed by appropriate semantic rules. Using these semantic rules, logical truths, and some trivial transformation rules, we can derive the conclusion that the relevant terms refer:

- (1)  $x$  is same color-ed as  $y$  iff the color of  $x$  = the color of  $y$ .
- (2) Beyoncé’s dress is same color-ed as Beyoncé’s dress.
- (3)  $\therefore$  The color of Beyoncé’s dress = the color of Beyoncé’s dress.
- (4)  $\therefore$  There is a color.
- (5) Colors are properties.
- (6)  $\therefore$  There is a property.
- (7) There is a property iff ‘property’ refers.
- (8)  $\therefore$  ‘Property’ refers.

On Thomasson’s view, every undischarged premise in this argument is either a semantic rule ((1), (5), (7)) or a trivial empirical truth ((2)).<sup>47</sup> In this fashion Thomasson can explain why metaphysical terms like ‘property’ refer: we have adopted terms governed by appropriate semantic rules, such that it follows, by trivial inferences and logical truths, that these terms refer. If it is objected that this does not explain why our term ‘property’ latches onto the properties ‘out there’ Thomasson can reply that this assumes a false view about the relation between the semantic rules of our language and the referents of our terms. The existence of these objects is not part of the explanation of why our terms refer. Instead, it is a trivial consequence of their reference. The reference of these terms is sufficiently guaranteed by their syntactic form (they are singular terms) and the fact that their introduction into

---

<sup>47</sup> It is empirical because it entails that Beyoncé’s dress exists, which, presumably, is not something one can know a priori.

our language does not produce any contradictions or inconsistencies with known empirical truths.

There is no further sense to the questions ‘do they really exist?’ or ‘do we really refer to them?’

I take it that other neo-Carnapian meta-ontologies can make similar moves. Whatever other problems there are with Thomasson and other neo-Carnapian views, they do not face the explanatory gap that Kant pointed out.<sup>48</sup> In terms of the argument given in the previous section, this means that the first horn of the dilemma (‘(i) substantivism is true, but there is no explanation of why metaphysical terms refer’) puts pressure specifically on substantivism. The Kantian objection I am articulating applies only to substantive meta-metaphysical views.<sup>49</sup> Nor does it apply to ‘descriptive metaphysics’ of the kind pioneered by Strawson 1966, which aims to describe our conceptual scheme. In the remainder of this paper, ‘metaphysics’ and ‘metaphysicians’ should be understood as appropriately restricted.

#### §4. Dismissing the Explanatory Demand

Before continuing I want to amend slightly P3 from Section 2, for the metaphysician can explain why *some* metaphysical terms refer. For instance, if ‘extended’ refers to the property of being extended and ‘simples’ refers to simples (objects with no parts), the metaphysician has an explanation why ‘extended simples’ refers to extended simples, if there are any.<sup>50</sup> However, the reference of ‘extended simples’ to extended simples piggybacks on the reference of ‘extended’ and ‘simples’. So the metaphysician has only explained how one term refers by assuming that others do. And, what is more, this explanation was only possible because the meaning of the one metaphysical term (‘extended simple’) is given in terms of two other terms (‘extended’ and ‘simple’). The moral of this example is that, when we can define one term via others, we can explain the reference of the defined

---

<sup>48</sup> Which does not mean they are unproblematic by Kantian lights. I criticize neo-Carnapian views in my in-progress monograph, *How is Metaphysics Possible?*

<sup>49</sup> However, Kant also thinks the Humean empiricist cannot explain why metaphysical concepts (especially those, like <cause>, that are concepts of an objectively necessary connection) refer, and this objection can be transformed into an objection to Carnap-inspired forms of empiricism, but that is outside the scope of this essay.

<sup>50</sup> See McDaniel 2007.

terms using the reference of the terms from which it is defined. This means that we should reformulate P3 as:

(P3\*) If substantivism in meta-metaphysics is true, there is no explanation of why primitive metaphysical terms refer, either to objects in the world or to worldly structure.<sup>51</sup>

The primitive terms of a metaphysical theory are the terms that are not defined in terms of other terms. For instance, the primitive terms of David Lewis's metaphysics might be: parthood, the relation of an object to its singleton set (e.g.,  $a, \{a\}$ ),<sup>52</sup> spatiotemporality, and the relation of objective natural similarity. It might be that the terms in a metaphysical theory are not explicitly defined in terms of more primitive terms, but that the whole theory serves as a kind of implicit definition of all of them collectively. That would entail that we must raise the question of what explains their reference *collectively* rather than raising it about each term individually.

In this section, I will consider several strategies the metaphysician might deploy to resist or evade the demand for an explanation of the reference of the terms in her theory. In the next section I will consider substantive explanations she might offer, by applying extant theories of reference to the case of metaphysics.

1. *No distinctive vocabulary.* The metaphysician might claim that metaphysics has no distinctive terminology of its own; all of its questions and claims can be formulated in the language of some other discipline (e.g., logic, mathematics, physics). Consequently, there is no distinctive phenomenon of metaphysical reference to be explained; there are only local phenomena of reference in mathematics, physics, and logic (assuming logic is referential at all).<sup>53</sup> Sider 2001 points out, for instance, that certain ontological questions ('how many objects exists?') can be formulated using purely logical terminology: quantification, identity, negation, and conjunction. But it is implausible

---

<sup>51</sup> Thus, Kant raises the question about the objective validity of the categories, which are the 'elementary' (A64/B89) or 'ancestral' concepts of the pure understanding rather than the 'derivative' (A64/B89, A81/B107) ones. An explanation of the objective validity of the derivative concepts can piggy-back on an explanation of the primitive ones, but not vice versa.

<sup>52</sup> Lewis 1991 attempts to build all of set theory out of mereology plus the primitive relation of an individual to its singleton-set.

<sup>53</sup> For contrasting views, see Wittgenstein 1922 and Williamson 2013.

that *all* metaphysical questions are like this. The recent literature on grounding, for instance, is clearly not about a purely logical notion, for grounding is not logical entailment. Nor, to use an example closer to Sider's heart, is 'structure' a term used in other fields (or at least it is not used to pick out the same notion as Sider's).

2. *Generalizations of ordinary vocabulary.* The metaphysician might retreat to a slightly weaker but related claim: terms as used in metaphysics (e.g., 'possible' and 'property') are simply generalizations of how these terms are used in the sciences and in everyday life. For instance, I might judge that it is possible it will rain today, and the chemist might judge that it is possible for hydrogen to be ionized.

Metaphysicians' talk of what is 'possible' is simply a generalization of this talk, in which various restrictions are removed (e.g., we are not talking only about what is likely to occur, or what is possible given the laws of nature). To put the point in terms of possible worlds, in metaphysics we speak with our modal operators 'wide open': we consider every metaphysically possible world, not merely those that are similar to the actual world. Similarly, we talk about properties in everyday life and in science, but in metaphysics we talk about *all* properties *whatsoever*.<sup>54</sup>

But if the ordinary notion is genuinely a restriction of the metaphysical notion, then, even if the ordinary term refers, it does not follow that the unrestricted metaphysical term refers. Consequently, an explanation of the first is not thereby an explanation of the latter.

This is easiest to see in the case of quantifiers.<sup>55</sup> Let us assume that ordinary non-metaphysical discourse uses a restricted quantifier  $\forall^R$  that takes as values of its bound variables only objects within a certain domain D (call it the *ordinary* domain) and that in metaphysics we talk with our quantifiers 'wide open', using an unrestricted quantifier  $\forall^U$  which we stipulate to be more general than  $\forall^R$  (so

---

<sup>54</sup> It remains to be seen whether some strategy like this could be applied to terms like 'grounding' and 'structure' which *prima facie* do not seem to be merely unrestricted forms of terms that are found in other contexts. I will remain neutral on this, however. Thanks to an anonymous referee for drawing my attention to this point.

<sup>55</sup> It will thus apply to talk about metaphysical modality if (a) we understand modal notions as quantifiers over possible worlds, and (b) we take metaphysical modal notions to quantify over a less restricted domain of worlds than ordinary modal notions. See the next note for a discussion of technical issues that arise for modality.

that  $\forall^U x Fx$  logically entails  $\forall^R x Fx$  but not vice versa). But if the ordinary quantifier ‘refers’ to (takes as values for its bound variables objects in) the ordinary domain, this by itself does not entail that the unrestricted quantifier ‘refers’ to (takes as values for its bound variables objects in) any larger domain. In particular, it does not entail that the unrestricted quantifier has a less restricted meaning (can take as a value of one of its bound variables an object not in D) even if there is an absolutely unrestricted meaning for it to have (if the domain of the restricted quantifier is not everything there is). We are owed an explanation of why our stipulation that  $\forall^U$  will be absolutely unrestricted (or merely less restricted) succeeds in conferring on it an absolutely unrestricted (or merely less restricted) meaning.

Likewise, physicists might talk about physical properties (e.g., mass, etc.), chemists about chemical properties, biologists about biological properties, etc. (call of these properties *scientific* properties) and then the metaphysician comes along, notices the commonality among these notions in different sciences, and wants to talk about *all properties whatsoever*. He could introduce the predicate ‘is a *property*’ by stipulating that it is a disjunction (i.e., it is equivalent to ‘is a physical property, or a chemical property, or a biological property, etc.’). But this makes it quasi-analytic the sentence ‘every property is either a physical property, or a chemical property, etc’ which indicates we have given ‘is a property’ the wrong meaning for metaphysics. For surely it should be an open question requiring substantive metaphysical research whether every property is physical or chemical or biological, etc. But if it is not stipulated to be disjunctive in this fashion, we are owed an explanation of why it refers to *all* of the properties there are (even in worlds where not all properties are scientific).<sup>56</sup>

---

<sup>56</sup> Williamson 2016 defines metaphysical necessity as the conjunction of all of the ‘objective’ necessities (e.g., nomic necessity). This might be thought to undermine my argument: if the objective necessities are in good shape semantically, then so is metaphysical necessity (by Williamson’s argument). But Williamson’s argument works entirely at the level of modal operators, understood as mappings from propositions to propositions (p. 456), and thus says nothing about the expressive resources of our language. It thus leaves unexplained why the modal idioms of our language are able to express all of the objective necessities in the first place. For instance, if the term ‘objective modal operators’ in English (or any term synonymous with it) picks out only a proper subset of the objective modal operators (set O, p. 456), then if we define metaphysical necessity as the conjunction of all of the necessity operators, using English, the defined operator might not express metaphysical necessity, but a more restricted necessity. Bacon 2018 also defines ‘the broadest necessity’ via quantification over ‘weak’ necessity operators, and thus fails to explain why our language has the

I anticipate that the substantive metaphysician will respond that all of these scenarios involve a highly ‘unnatural’ or gerrymandered reference relation, i.e., that the intended interpretation (absolutely unrestricted generality, all properties, the most general necessity) is more natural than the alternatives. This brings in the idea of naturalness as an explanation of reference, to which I return below in Section 6.

3. *Explanation by division.* The metaphysician might seek to dismiss the explanatory demand by distinguishing words in our language from their meanings. To take a toy model we might distinguish between our word ‘property’ and its meaning (sense), understood as an abstract Fregean concept ( $\langle \textit{property} \rangle$ ) that exists independently of our thought and talk. The fact that ‘property’ refers to properties decomposes into two distinct facts: (*p*) the fact that ‘property’ has the Fregean concept  $\langle \textit{property} \rangle$  as its sense and (*q*) the fact that this concept  $\langle \textit{property} \rangle$  refers to the set of properties or the property of being a property (or something like that). The explanation of the former is purely psychological/linguistic; we came to speak a language one of whose terms has  $\langle \textit{property} \rangle$  for its sense. The explanation of the other fact might be that Fregean concepts have their referents necessarily; it is part of the nature of  $\langle \textit{property} \rangle$  to refer to properties (the set of properties, the property of being a property, etc.).<sup>57</sup> Since the original fact (that ‘property’ refers to properties) divides without remainder into these two facts, and each of these facts has its own explanation, the original fact is explained. The point of this ‘explanation by division’ is not that there is nothing that needs to be explained here (e.g., how do terms in our language express Fregean concepts? what determines the reference of Fregean concepts?) but that what needs to be explained is not what we originally thought, so the ‘problem’ of metaphysical reference disappears.

---

semantic resources to express this broadest necessity; we might, after all, lack the resources to quantify over *all* weak necessity operators. Thanks to an anonymous referee for pressing me on this point.

<sup>57</sup> I am not endorsing the claim that  $\langle \textit{property} \rangle$  relates to its referents essentially. I am exploring this as an option that the metaphysician might pursue.

The conjunction of  $p$  and  $q$  entails the fact,  $r$ , that one and the same thing ( $\langle \textit{property} \rangle$ ) is the meaning of ‘property’ and refers to, say, the set of all properties (or the property of being a property, or something of that nature).<sup>58</sup> However, giving separate and unrelated explanations of the conjuncts leave unexplained their connection: the psychological explanation of  $p$  fails to explain why we come to express metaphysical concepts that refer, rather than non-referring metaphysical concepts. Compare: we might offer a biological-evolutionary explanation of why we hold the moral beliefs that we do (e.g., that incest is wrong) and a purely normative-ethical explanation of why those beliefs are true (e.g., a proof that incest is wrong from some more basic moral principle), but we would not thereby have explained why we hold true ethical beliefs rather than false ones. Intuitively, the reason we would have failed to explain this, is that we have not explained why the process by which we form our ethical beliefs (in this toy case, evolution) produces truths. In particular, we have not shown that there is a connection between why we form these beliefs and why they are true that would explain why the beliefs we form by the process cited in the explanation are truth-tracking. Likewise, ‘the explanation by division’ strategy fails to explain why the semantic/linguistic/psychological/what-have-you explanation of why we express the concepts we do ( $p$ ) is a process that results in our expressing referring concepts ( $q$ ). We have not connected our explanation of why we express the concepts we do with our explanation of why they refer, in such a way as to have explained why we express referring concepts in metaphysics. This strategy for dismissing the explanatory question about metaphysical term reference thus fails.

4. *Abductive reasoning.* Some defenders of metaphysics might object that this explanatory question incorrectly treats our metaphysical terms statically, as though we have a fixed stock of them that never changes. Instead, the metaphysician might argue, we begin with a set of metaphysical terms, either the deliverances of basic reflection on the world or some historical tradition (likely both), and the current state of our physical and other natural-scientific theories, and we try to produce the best overall explanation of the world. In the process our basic metaphysical terms change, from, say the

---

<sup>58</sup>  $r = \exists x(\text{‘substance’ means } x \ \& \ x \text{ refers to substance}).$

terms in Aristotle's metaphysics (substance, accident, activity, potentiality, etc.) to the terms of contemporary Lewisian metaphysics (part, set, abstract/concrete, etc.). Departing from the linguistic focus of the rest of this essay, this point is best appreciated in terms of concepts rather than linguistic terms, to make clear that we are individuating terms semantically not syntactically. It is obvious that the terms used in metaphysics have changed syntactically since Aristotle (e.g., from 'οὐσία' to '*substantia*' to 'substance'); the important point is that the terms we now use in metaphysics mean different things than Aristotle's terms, i.e., they express different concepts.

Distinguish two questions. First, an epistemic question: what reason do we have for thinking that our metaphysical concepts refer? Secondly: assuming they do refer, what is the explanation of this fact? I have been pushing the second question, but the abductive conception of metaphysics is at most an answer to the first, epistemic question. It provides us, at most, a reason to think that our concepts do refer *unless* an explanation is given of why *in general* the terms of abductively well-supported theories refer. Without such an explanation, the abductive picture of metaphysics might provide us all the reason in the world to believe that our metaphysical concepts refer without doing anything to *explain* this putative fact. It will not, in other words, answer Kant's question.

Does the metaphysician possess an explanation of why the terms in an (even ideally) abductively justified metaphysical theory refer? There are two moves she might make. The first is a metaphysical version of the 'no miracles' argument in the philosophy of science.<sup>59</sup> Just as the empirical predictive success of our scientific theories would be miraculous if their terms did not refer (or so the scientific realist claims), the explanatory success of our metaphysical theories would be miraculous if their terms did not refer. But, again, this is only an answer to the epistemic question: a reason to think that the terms in our metaphysical theories refer, not an explanation of why they do. What is more, far from a history of stunning empirical successes, the history of metaphysics is, arguably, a history of wreckage (albeit glorious, fascinating wreckage). A 'no miracles' style argument is neither available nor dialectically helpful to the metaphysician at this point. The second thing the metaphysician might

---

<sup>59</sup> Originally adumbrated by Putnam 1975*a*, 73.



try is to pass the buck to the natural sciences: whatever explains the reference of the terms in our best scientific theories explains the reference of our metaphysical terms. In the next two sections I will argue that none of the extant explanations of why natural scientific terms refer will help the metaphysician, so we are back where we started: without an explanation of why metaphysical terms refer.

### §5. Applying Extant Theories

In this section I consider whether various extant theories can be marshaled to explain the reference of terms in metaphysical theories. I argue, in each case, that extant theories of reference are of little help to the metaphysician.

*Causal.* Causal theories of reference, very broadly, hold that reference is explained by causal relations between tokenings of expressions and their referents.<sup>60</sup> To take a toy model, ‘water’ refers to water because initial tokening of ‘water’ are causally related in the right way to (samples of) water (while later tokenings inherit their reference by historical chains of transmission). I take it to be relatively clear that causal theories of reference will not be of much help in explaining metaphysical term reference. This is not because I am assuming that metaphysics concerns itself exclusively with abstracta, which are (at least standardly taken to be) causally inert. As I mentioned in Section One I am focusing primarily on metaphysical theories about concrete reality, in part because I want to show that the problem of metaphysical term reference is not merely the problem of referring to abstracta.

Aside from the numerous problems that beset causal theories in general<sup>61</sup>, the reason that causal theories of reference will not help explain the reference of metaphysical terms (even theories about concrete reality) is that, while the topic of a metaphysical theory might be concrete reality endowed with causal powers, metaphysics typically considers its object at a level of abstraction above that at which its causal powers can be located. Let us take what is perhaps the best case for the causal theorist: properties. On many views of causation, properties play an important role in causal

---

<sup>60</sup> E.g., Kripke 1980, Devitt 1981.

<sup>61</sup> See the summary of these problems in Adams & Aizawa 2017.

relations<sup>62</sup>; indeed on some views, they are the principal causal relata.<sup>63</sup> But while, for instance, the acidity of the solution (understood as an event, a property-exemplification, a trope, or what have you) may cause the litmus paper to turn red, *that acidity is a property* is not part of the causal explanation of the paper's changing color. If it were, the correct causal explanation of the change would have to include the fact that acidity is a property. The complete *physical* explanations of an event (e.g., the tokening of 'that is an acid') will presumably cite properties (e.g., the property *being an acid*), but it will not cite the 2<sup>nd</sup> order property *being a property*. This means that *being a property* does not stand in cause/effect relations, thus is not a causally relevant factor in tokenings of words. The fact that a sample has the property of being water is part of the causal explanation of tokenings of 'water' (according to the causal reference theorist), but the fact that *being water* is a property is not. In general, since the causal theorist explains the reference of terms by citing the causal factors that influence initial tokenings (which reference is then transmitted to later tokenings by historical chains of transmission) this means that the casual theorist may be able to explain why we refer to various objects, properties, kinds, etc. but not why we refer to the properties *being an object*, *being a property*, *being a kind*, etc. This problem ramifies into any area of metaphysical inquiry where the objects are constituents of concrete reality, but the terms in our theories refer to features and structures that do not figure in causal explanations. For instance, while concrete objects have various *de re* modal properties (e.g., they are possibly one way and not possibly another) that something has modal features is, quite plausibly, not part of the causal explanation of our tokenings of various modal expressions.<sup>64</sup>

---

<sup>62</sup> E.g., the classic analysis of Kim 1973, which takes causation to be a relation among events, where an event is the exemplification of a property by an object at a time (e.g., the event of the stone being warm at *t*). Cf. Lewis 1973.

<sup>63</sup> Campbell 1990.

<sup>64</sup> Though its possession of those *de re* modal properties may be explanatorily related to its causal powers, e.g., it is *de re* necessarily water because it is water (and everything that is water is essentially water, etc.). Likewise, the fact that *acidity* is a property is explanatorily related to the causal power that samples of acidity have, because causation is a relation between properties or property-exemplifications (events), let us assume. But this explanatory connection is not a *causal* one. Even though acidity-exemplifications are causes partly in virtue of the fact that acidity is a property, that acidity is a property does not *cause* acidity-exemplifications to be causes. But the causal theory of

Friends of the causal theory of reference might reply that this merely requires a more fine-grained understanding of the causal relation between the world and tokenings of metaphysical expressions like ‘property’ and ‘possibly,’ one that includes factors like being a property and having *de re* modal features. But working this out would require a whole theory of causation that integrates these factors into a causal explanation of ordinary events like word-tokenings. To say that developing such a causal theory of reference would be an enormous undertaking would be an understatement: it would require showing how every part of our metaphysics plays a causal role in the original tokening of the term in the theory that refers to it. This does not seem like a promising strategy for defending metaphysics. No wonder, then, that no extant causal theory of reference has attempted it.<sup>65</sup>

*Externalist.* Externalist views of reference hold that the facts that determine the referent of an expression do not supervene on the internal states of speakers who understand the term; consequently, holding fixed the internal states of speakers, as we vary the environment in which the term was first tokened, we potentially vary the referent of the term. To use the classic example from Putnam 1975*b*, if ‘water’ is first tokened on Earth it refers to H<sub>2</sub>O, but if we consider it as first tokened on Twin Earth it refers to XYZ.

This negative formulation of externalism is intended to bring out the fact that externalism is not a kind of explanation of reference-facts, but a *phenomenon* that is explained by certain theories of reference. For instance, the causal theory that the referent of a term like ‘water’ is whatever natural kind was causally relevant to initial tokenings of the expression entails that the referent of ‘water’

---

reference is a theory specifically about *causal* determination, not determination in general: the referent of a term is what *causally* determines its tokenings. In response to this problem, one might broaden the causal theory and allow as referents of terms entities that play an explanatory role (don’t necessarily causal) in the tokening of those terms. But this has the effect of making reference too easy, for assuming the world is explanatorily well-founded (there is a fundamental level that explains everything else), just by using language I will have thereby referred to the metaphysical fundamental level. Metaphysics may be possible (I certainly believe it is), but it is not that easy.

<sup>65</sup> Simchen 2017 distinguishes between ‘interpretationist’ and ‘productivist’ theories in meta-semantics. Aside from the causal theory of reference, the theories of reference I discuss in this essay are largely interpretationist. This is because, aside from the causal theory of reference for natural kind terms (which, I argue in the body of the paper, cannot be extended to metaphysical terms), the focus in productivist theories has been on expressions that do not play much role in metaphysics (e.g., demonstratives, proper names).

does not supervene on the internal states of speakers but partly depends on which environment (whether Earth or Twin Earth) the term was originally tokened in. Consequently, externalism as such does not constitute a candidate explanation of metaphysical term reference, so we must turn our attention to specific families of such explanations.

*Descriptive.* A descriptive theory of reference for a given class of expressions associates with each such expression a description, and holds that the referent of the expression is that object or objects that uniquely satisfy that description. Descriptive theories will differ in how they associate terms with descriptions (e.g., whether a competent speaker must be able, upon reflection, to specify the description associated with the term), but those details will not matter for our purposes here.

I think we can dispense with a descriptive explanation of metaphysical term reference fairly quickly. The terms whose reference I have been inquiring into are the *basic* terms in a metaphysical theory, the ones that cannot be explicitly defined in more basic terms. What then can serve as the reference-fixing definite descriptions for those terms? The rest of the theory provides a kind of implicit definition of its terms: the terms refer to whatever they must to make the theory come out true. The trouble with this kind of global descriptivism, as Hilary Putnam (following Quine) pointed out, is that if the theory is consistent then provably there are multiple models that satisfy the theory.<sup>66</sup> Global descriptivism underdetermines reference. The problems with descriptivism have inspired some metaphysicians to include a notion of ‘naturalness’ or ‘joint carving’ in their theories; I discuss them in the next section.

*Knowledge-maximization.* Before going into depth about naturalness, though, I want to discuss the ‘knowledge maximization’ view of Williamson 2007. While Williamson agrees that naturalness plays

---

<sup>66</sup> The model-theoretic argument first appeared in Putnam 1977, then in a more formal version in 1980, and then again in numerous publications throughout Putnam’s career. I cite Putnam 1981, 22–48, because it is the most widely read version. I ignore later formulations of the argument because Lewis 1984 (and following him, Sider 2011) only discusses these initial three. I also forego a formal presentation of the argument; interested readers should consult those Putnam texts. Lewis 1984 and Sider 2011, 28–32, both point out that the argument does not depend on the Löwenheim-Skolem theorem (as Putnam claims), but, instead, on some simple considerations of elementary model theory.

some role in fixing reference, he thinks that naturalness by itself is not sufficient to plug the holes in global descriptivism. Instead, Williamson proposes an epistemic constraint on interpretations of speakers' referents: the correct interpretation (assignment of referents) is the one that maximizes knowledge. Williamson does not analyze or define knowledge, for he adopts a 'knowledge first' approach: knowledge is the most general factive mental state, the most basic way in which our minds are related to reality.<sup>67</sup> Williamson's proposal is that among all of the 'deviant' interpretations of our language that Putnam's argument generates, we can eliminate those that assign to our terms semantic values with which we could not be in epistemic contact. Our interpretation should maximize not only truth and naturalness, but knowledge as well.

Even if Williamson is correct that this determines a unique, or relatively compact, set of assignments for the terms in our language, his 'knowledge maximization' view is not dialectically well placed to answer the Kantian question about metaphysical term reference. Williamson's view applies to our language *as a whole*.<sup>68</sup> But metaphysics is a relatively small part of our language, so the fact that the correct interpretation of our language as a whole maximizes knowledge gives us little hope that the one corner of it that constitutes metaphysics does as well. Alternatively, Williamson could apply the view at the level of individual theories, but that would be problematic in its own right: why assume that a given theory refers at all, why assume that it is in semantic contact with reality? Take for instance, demonology and metaphysics. If we apply Williamson's 'knowledge maximization' principle to metaphysics (rather than to our whole language), why should we not apply it to demonology as well? The natural answer is that that metaphysics is on better semantic footing because it is on better epistemic footing: the methodology of metaphysics (e.g., abduction) is continuous with the methodology of the sciences, according to Williamson. This brings me to the final respect in which Williamson's view is not dialectically well placed to answer my question. By

---

<sup>67</sup> Williamson 2007, 269; for a more comprehensive defense of the 'knowledge first' approach, see Williamson 2000.

<sup>68</sup> He does so in order to avoid otherwise intuitively compelling counterexamples; see Williamson 2007, 267.

applying the ‘knowledge maximization’ principle, not to our whole language, but to metaphysics in particular, Williamson can explain the reference of our metaphysical terms only on the assumption that metaphysics is within our epistemic grasp, i.e., that we can have knowledge in metaphysics (e.g., by abduction). But this turns the question about the semantic possibility of metaphysics (why do metaphysical terms refer?) into a question about the epistemic possibility of metaphysics (why do we have knowledge in metaphysics?), but skeptics about the former were already skeptics about the latter (how can we have metaphysical knowledge if metaphysical terms do not even refer?), so it is not clear how much dialectical progress has been made. To be sure, Williamson may be able to explain the possibility of metaphysical knowledge (or overcome skeptical objections to its possibility) and thereby explain metaphysical reference, via the ‘knowledge maximization’ view of reference, but deciding this lies outside the scope of this essay.<sup>69</sup>

### §6. Joint Carving and the Possibility of Metaphysics

Lewis 1983 and 1984 answered Putnam’s model-theoretic argument against global descriptivism by introducing a primitive notion of ‘naturalness,’ a notion that was taken up and further developed by others, most notably by the account of ‘joint-carving’ and ‘structure’ in Sider 2011.<sup>70</sup> My account of structure will draw heavily on Sider, but it might be worth taking a step back from the specifics of Sider’s theory and considering the idea of joint carving in general. The basic idea is that the world has an objective structure. It has this structure independently of how, and whether, we think of it at all. Hence the metaphor of carving at the joints: just as we can butcher an animal carcass at the joints (the original source of Plato’s metaphor<sup>71</sup>) or fail to do so (e.g., by quartering it), we can ‘carve’ the

---

<sup>69</sup> Williamson might respond that the ubiquity of terms like ‘property’ and ‘causation’ in discourse outside metaphysics makes it overwhelmingly likely that they refer, via the knowledge-maximization principle (for otherwise our discourse in those areas would not constitute knowledge). Even if this strategy succeeds, it will not easily be extended to more recherché topics of metaphysics, e.g., grounding, possible worlds, or instantaneous temporal parts.

<sup>70</sup> Williams 2007 constructs a set of counterexamples to Lewis’s view, which, arguably, also constitute counter-examples to Sider’s. My objections to Sider, however, are pitched at a greater level of generality; even if Sider had a reply to Williams’ counterexamples, my critique would still apply. Cf. Hawthorne 2006, 58–61, for further discussion of naturalness and reference magnetism, and Weatherson 2003 for some objections to the standard reading of Lewis.

<sup>71</sup> *Phaedrus*, 265e-266a.

world (i.e., use terms that refer to) at its ‘joints’ (structural aspects of the world) or fail to do so (describe the world in a way that does not correspond to its objective structure).

Joint carving, or ‘structure’ as he calls it, plays an important role in Sider’s theory of reference.<sup>72</sup> The ‘descriptive content’ of a term is specified by some subset of the true sentences of the language containing that term: those that are analytic, those that are taken to be meaning-constitutive, or what have you. The details will not matter, so long as they are a subset of the true sentences of the language.<sup>73</sup> The descriptive content of a term determines a range of ‘reference candidates,’ those that make that descriptive content true. For instance, the descriptive content of ‘water’ determines a range of reference candidates, each of which is wet, found in lakes and rivers, etc.<sup>74</sup> The actual referent of the term is the reference-candidate that is maximally structural (in the case of ‘water,’ H<sub>2</sub>O). If there is no such unique candidate (if multiple candidates are equally structural) then there is no determinate fact as to which one the term refers to. In this case, there are questions involving the term that are said to be ‘non-substantive.’ For instance, because the properties *unmarried male* and *unmarried male eligible for marriage* are equally structural, the question of whether the pope is married is non-substantive (its answer depends on which of two equally structural interpretations of ‘bachelor’ we adopt). Let us call such a term, for which there are multiple equally structural interpretations, and thus about which there will be non-substantive questions, an ‘*indeterminate*’ term.<sup>75</sup>

Can Sider’s theory of structure explain why metaphysical terms refer? Take some metaphysical term like ‘particular’ and its descriptive content, e.g., ‘particulars instantiate properties but are not

---

<sup>72</sup> Sider 2011 prefers ‘structure’ because he does not want to be ontologically committed to joints. However, I will use ‘joint-carving’ and ‘structure’ interchangeably; this should not be read so as to incur ontological commitment to joints.

<sup>73</sup> Otherwise, every truth will be true in virtue of the reference-fixing truths (because it is one of them), which is, intuitively, absurd.

<sup>74</sup> We can relax the requirement that the descriptive content must all be true, but that complication will not crucially matter here. For instance, in the case of natural kind terms we might discover that the descriptive content of the term is false of its referent (e.g., in the early modern period we discovered that water is not an element.)

<sup>75</sup> I am assuming throughout that, if there are multiple equally structural reference-candidates for a term, then there will be at least one non-substantive question about that term, i.e., the question of whether that term is one of the equally structural reference-candidates (e.g., whether bachelors are adult unmarried males eligible for marriage).

themselves instantiated.’ Assume there is an absolutely structural property, the property of instantiating properties but not being instantiated. The term ‘particular’ refers to this property because the property is (i) a reference-candidate for the term (it fits the descriptive meaning, assuming ‘instantiation’ refers to instantiation, etc.) and (ii) it is maximally structural.<sup>76</sup>

My question is: *why* do the terms in our metaphysical theories refer to structural items? At several points, Sider argues *that* reference in general is structural, but I will largely set aside these arguments,<sup>77</sup> for I am not interested in *whether* reference is structural; I am interested in *why*. Sider’s best response, I think, is to reject my request for explanation as confused: it is simply part of the essence of reference, part of what it is to *be* reference, that it maps terms in our language to structural items. My request for an explanation of why we refer to structural items is no more coherent than the request for an explanation of why 1 is the successor of 0, or why water contains oxygen: that is just part of what it is to be 1 or to be water. While I am suspicious of the dialectical appropriateness of Sider simply building structuralness into the very nature of reference (it feels like cheating), I will temporarily grant him that maneuver and argue that his theory of metaphysical reference still faces serious difficulties.

Sider’s official notion of structure is absolute: a term is either structural or not structural at all.<sup>78</sup> But his view cannot be that it is part of the essence of reference that we refer only to absolutely structural items, for that would entail, absurdly, that we could only ever talk about the absolutely

---

<sup>76</sup> Theories like Lewis’s and Sider’s are sometimes referred to under the rubric ‘reference magnetism.’ But I take the metaphor of magnetism to just mean that *ceteris paribus* the referents of our terms are ‘natural’ (Lewis) or ‘structural’ (Sider), so I will largely dispense with this metaphor in what follows.

<sup>77</sup> There is the argument from semantic explanation (Sider 2011, p. 28) and the argument from stipulation (p. 74). The former is clearly not intended as an explanation of why reference is structural, but only an argument that it must be (semantic facts are explanatory, and explanations must be comparatively structural). The latter might seem to be more of an explanation (we stipulate that henceforth our language is to be interpreted in the most structural way possible), but it presupposes that we can refer to the structural (otherwise the stipulation will fail), so I think it is secondary to the explanation discussed in the main text (that reference is essentially structural.) Sider addresses this at one point (p. 74) and argues that we must be able to bootstrap some new theoretical terms, but he does not explain why bootstrapping ‘structure’ can succeed, so I pass over those arguments here. Finally, Sider could argue that reference itself is a ‘reference magnet’; but I think that argument relies on the assumption that referents are *ceteris paribus* structural (hence ‘reference’ refers to reference, rather than a less structural reference relation), i.e., that reference is essentially structural. This is why I focus on that claim in the main text.

<sup>78</sup> Sider 2011, p. 128.



structural items but never about ordinary objects (which are not absolutely structural). It would also entail that we could never refer to reference itself (because reference, being a relation between something less than absolutely structural, namely our language, and the world, is itself less than absolutely structural), rendering the view semantically self-undermining.<sup>79</sup>

Sider thus needs a comparative notion of structure: one item can be more structural than another, even though neither is absolutely structural. At one point Sider considers the following account of comparative structuralness: T is at least as structural as T\* just in case the ‘metaphysical definition’ of T in a perfectly structural language is at least as short as that of T\*.<sup>80</sup> (What exactly a metaphysical definition is will depend on the details of Sider’s preferred language for the ‘book of the world.’)<sup>81</sup> But Sider ultimately rejects this account, for it entails that the disjunction of two perfectly structural properties is just as structural as their conjunction. Since the structural is supposed to be more explanatory than the non-structural and, intuitively, conjunctions are more explanatory than disjunctions, an account merely in terms of the length of definitions cannot be correct. But Sider does not replace it with an explicit account of relative structuralness that resolves these problems.<sup>82</sup>

That reference is structural could mean at least two different things (*a*) the reference relation itself (between terms in our language and their assignments on the intended interpretation) is comparatively structural (i.e., more structural than other reference-like relations, i.e., other interpretations of our language), or (*b*) the referents of terms in our language are relatively structural (i.e., more structural than those assigned by other interpretations). I will focus on (*b*) because it is the

---

<sup>79</sup> Below, I consider whether Sider could retreat to a position on which we can only refer to the absolutely structural. For now, I will just treat it as a dialectical dead-end.

<sup>80</sup> Sider 2011, p. 129–133.

<sup>81</sup> For instance, the metaphysical definition of a singular term might be a definite description in purely structural terms that can be substituted *salva veritate* in all sentences that contain that term. Likewise, the metaphysical definition of a predicate might be an open sentence in purely structural terms that can be substituted *salva veritate* in all sentences that contain that predicate, etc.

<sup>82</sup> He suggests that this problem could be resolved by requiring all definitions to be in some standard form (e.g., prenex disjunctive normal form), but does not further develop this suggestion (Sider 2011, p. 131).

structuralness of the relata (the referents of our terms) that is more important for the dialectic than the structuralness of the reference relation itself.<sup>83</sup>

In some discussions of ‘reference magnetism’ Sider explains the reference of individual terms while taking for granted the background meaning of the other terms in our language. For instance, given that ‘water is the clear drinkable fluid found in lakes and rivers’ has to come out true (it partly specifies the descriptive content of ‘water’), and given that ‘lakes’ refers to lakes, etc., ‘water’ refers to H<sub>2</sub>O because it fits the descriptive meaning of water (it is found in lakes, etc.) and it is more structural than the other reference-candidates (let us suppose).<sup>84</sup> But in doing so we assume background semantic facts about the other terms in our language, specifically, that they have reasonably structural meanings (e.g., that ‘liquid’ refers to liquid). Therefore, this kind of explanation does not explain why our language *as a whole* has such comparatively structural meanings. To explain that fact we need to take our language as a whole and compare reference to different global interpretations of that language.<sup>85</sup>

Before continuing, let me be explicit about how I am using some of this terminology. I am treating reference as an interpretation function on the sentences and sub-sentential expressions of our language, akin to an assignment function in model theory (i.e., in a first order theory, one that assigns extensions to predicates, objects to singular terms, truth-values to sentences, etc.). Other ‘interpretations’ are just other such assignment functions in other models. When I talk about the

---

<sup>83</sup> If the structuralness of non-fundamental relations supervenes on the structuralness of their relata then the structuralness of (*a*) entails the structuralness of (*b*). If we assume that reference is more structural than other reference-like relations, since we are holding fixed one set of relata (terms in our language), (*a*) entails that reference will assign more structural referents than the other interpretations. However, it is unlikely that Sider would hold that the structuralness of non-fundamental relations supervenes on the structuralness of their relata, for this would entail that every relation among non-fundamental items is equally structural (e.g., intuitively structural spatiotemporal relations among ordinary objects would be just as structural as arbitrarily gerrymandered ones.) But, assuming that he denies the supervenience claim, Sider has no obvious path from the claim that reference is a comparatively structural relation to the claim that its relata are; this is why I focus on (*b*) in the main text.

<sup>84</sup> Sider 2011, p. 46–7.

<sup>85</sup> As Sider does elsewhere (e.g., p. 23–35). I don’t take myself to be criticizing Sider here, but simply explaining why my own argument will consider the global interpretation of the language as a whole, rather than term-by-term interpretation.

‘referents’ of our terms, I mean the values they are assigned under the *reference* assignment function, the intended interpretation; other interpretations will be said to assign other ‘meanings’ to our terms.<sup>86</sup> When I talk about the ‘domain’ of an interpretation, I do not mean merely the domain the set of objects used to interpret singular terms and quantifiers, but everything that the interpretation assigns as the meanings of terms in the language, including objects, predicate-extensions, etc. This means that ‘domain’ (as I am using the term) refers to something that cross-cuts various logico-syntactic categories. But this is appropriate in the context of Sider’s theory, for structuralness applies to items of, potentially, any logical category. Objects, predicates, quantifiers, operators, etc. can all be evaluated with respect to their structuralness. We thus need a term of equal logico-syntactic generality to cover all the items that can be evaluated for structuralness on a given interpretation of our language. I have chosen ‘domain’ for this purpose. When I mean domain in the standard sense of the domain for interpreting the quantifiers (in standard model theory, a set), I write ‘quantificational domain.’ The reason this is important is that, in evaluating different interpretations in terms of the structuralness of their domains, I am not assuming that the structuralness of an interpretation consists solely in the structuralness of the objects in its quantificational domain (which Sider, rightly, would reject.)

Sider never provides an explicit metric on the structuralness of possible interpretations of our language, so it is somewhat difficult to evaluate the claim that reference is more structural than other interpretations. For instance, here are two possible metrics, where  $D$  and  $D^*$  are the domains of such interpretations (the items assigned as the meanings of our terms on those interpretations):

*Max-max.*  $D$  is at least as structural as  $D^*$  iff the most structural item in  $D$  is at least as structural as the most structural item in  $D^*$ .

*Min-max.*  $D$  is at least as structural  $D^*$  iff the least structural item in  $D$  is at least as structural as the least structural item in  $D^*$ .

---

<sup>86</sup> Earlier I talked about reference as a relation, but I take it we can easily translate between reference-as-relation  $R$  and reference-as-function  $r$ :  $R(x, y)$  if and only if  $r(x) = y$ .

But, clearly, neither of these metrics can be correct. In the case of *Max-max*, consider an interpretation in which the most structural items in the domain are as structural as the most structural items in the intended interpretation (reference), but where every other item is arbitrarily non-structural; this would be an interpretation that agrees with the intended one on all absolutely structural terms but disagrees everywhere else. *Max-max* rule entails it is just as good an interpretation structure-wise as reference. Sider must reject his, for he wants to use structure to explain why terms without absolutely structural meanings determinately refer to comparatively structural items, rather than being indeterminate.<sup>87</sup> But neither can the *Min-max* rule be correct; consider a model that agrees with the intended model on the least structural terms but assigns arbitrarily non-structural meanings to all others as well. The *Min-max* rule entails this is just as good an interpretation structure-wise as the intended interpretation.<sup>88</sup> This is clearly a no-go for Sider, who wants to explain why terms in fundamental metaphysics (absolutely structural metaphysics) determinately refer to absolutely structural items.<sup>89</sup>

If we cannot evaluate the structuralness of interpretations in terms of their most (*Max-max*) or their least (*Min-max*) structural items, then the comparative structuralness of interpretations must be a global matter:

*Sum:* D is at least as structural as D\* just in case the sum of the structuralness of the items in D is at least as great as the sum of the structuralness of the items in D\*.

This is merely a schematic stand-in for an actual principle, though, for it says nothing about what the summation function is, whether, and how, the structuralness of different terms is weighted, etc. But, setting these issues aside, we can ask: is there a unique maximally structural interpretation of our language? In fact, as we have seen, Sider's answer has to be 'no'. There are multiple, equally overall

---

<sup>87</sup> Assuming the different, equally structural, meanings differ in some properties, it would also massively over-generate non-substantive questions involving less-than-absolutely structural terms.

<sup>88</sup> Sider identifies the fundamental with the absolutely structural (Sider 2011, p. 115–6).

<sup>89</sup> My argument in this paragraph is not that there are such deviant interpretations, if Sider adopts Max-max or Min-max, but that if he adopts either metric he cannot rule out such deviant interpretations on structural grounds, and that is precisely the role of structure in his meta-semantics. See below for more on the shape of the dialectic here.

structural ways of interpreting our language. For instance, consider the interpretation of the whole of our language that takes the referent of ‘bachelor’ to be *unmarried male* and the one that agrees with it in all respects, except that it takes the referent of ‘bachelor’ to *unmarried male eligible for marriage*. These are equally structural reference candidates, so there is a tie in overall structuralness between the two interpretations. This is, so to speak, a feature not a bug of Sider’s theory.

However, if Sider abandons *Max-max* or *Min-max* in favor of something like *Sum*, there is a worry that his view will allow trade-offs in absolute structure. By ‘trade-offs in absolute structure’ I mean interpretations that are overall just as structural as the intended one (reference), but which assign less than absolutely structural meanings to some terms in fundamental metaphysics, while compensating by assigning more structural meanings to terms elsewhere in our language. Such trade-offs would pose a deep problem for fundamental metaphysics, for they would entail, with respect to the relevant terms, that we do not determinately refer to the absolutely structural: there would one interpretation on which some term in fundamental metaphysics (e.g., ‘spacetime point’) refers to something absolutely structural and an equally overall structural interpretation on which it refers to something less-than-absolutely structural. This would not only generate non-substantive questions in fundamental metaphysics but would make the claim that fundamental metaphysical is fundamental (is about the absolutely structural) itself non-substantive: there is one interpretation on which it is about the fundamental, and another equally structural interpretation on which it is not. This would be a serious problem for Sider’s conception of metaphysics.

In light of this, Sider might propose a fourth metric, on which ‘losses’ in absolute structure cannot be compensated by structural ‘gains’ elsewhere in an interpretation (where  $D$  and  $D^*$  are as above):

*No absolute trade-offs* (NATO):  $D$  is at least as structural as  $D^*$  iff (i) the sum of the structuralness of the items in  $D$  is at least as great as the sum of the structuralness of the items in  $D^*$  and (ii) the most structural items in  $D$  are at least as structural as the most structural items in  $D^*$ .

Since nothing is more structural than the absolutely structural, if D contains absolutely structural items, no interpretation D\* that contains only comparatively structural items will at least as structural as D; losses in absolute structuralness cannot be ‘compensated’ by gains in comparative structure elsewhere.

However, even NATO will not save Sider’s explanation of metaphysical reference. Where D is the intended domain (the domain of referents), another domain D\* might be overall as structural as D and contain absolutely structural items; by NATO it is at least as structural as D and thus structure-wise as good a global interpretation of our language as the intended one. But if D\* does not agree with D on what it assigns as the absolutely structural meanings of terms in fundamental metaphysics, this generates referential indeterminacy and non-substantive questions in fundamental metaphysics. It generates referential indeterminacy because the terms of fundamental metaphysics do not determinately have their D-meanings rather than their D\*-meanings. This in turn generates non-substantive questions, for, if their D- and D\*-meanings differ in any absolutely structural respects, this will mean there are questions in absolutely structural terms that have no determinate answers.<sup>90</sup> But now Sider’s fundamental metaphysics will be in no better position vis-à-vis the absolute structure of the world than Putnam’s ‘metaphysical realist’ (the original target of the permutation arguments) is vis-à-vis the world generally: it will be determinately the case that she is referring to something (Sider: something absolutely structural) and saying something true (Sider: something true and cast in in absolutely structural terms) about it, but it is indeterminate what she is referring to and what she is saying about it (in Sider’s case: what bit of absolutely structure she is talking about and what absolutely structural truth about it she is saying.) Since Sider introduces structure into meta-semantics

---

<sup>90</sup> I take it that Sider can allow some indeterminacy at the absolutely structural level, i.e., quantum indeterminacy about the position and velocity of fundamental particles. However, the kind of indeterminacy I am considering in the main text is significantly worse for Sider than this: it would arise even in a completely determinate world and would arise purely meta-semantically. The kind of indeterminacy I am considering is the kind of indeterminacy Putnam argued was a consequence of ‘metaphysical realism’ (whatever exactly that is) and that Sider’s whole meta-semantic theory is designed to ward off.

precisely to avoid this kind of referential indeterminacy (like Lewis 1984 did with primitive naturalness), I take this to be a serious problem for his account.

To all of this Sider can respond: this would be a problem if there were such an interpretation (one that is at least as structural as reference by NATO but differs in its interpretation of terms in absolute metaphysics) but there isn't one, and no argument has been given that there is one. But my point is that Sider has not shown that *there is not* one and, insofar as he not done this, he has failed to explain satisfactorily why determinate reference in fundamental metaphysics is possible. He has explained why determinate reference in fundamental metaphysics is possible on the assumption that there no such deviant-but-equally-overall structural interpretations, but this assumption stands at least as much in need of explanation as anything else. More specifically, his explanation presupposes that the metric of structuralness over the space of possible interpretations has a family of maxima that do not differ in their interpretations of absolutely structural terms. Since he does not provide a worked-out metric on structuralness he is no position to explain why this is the case. An explanation that itself rests on a such a highly contentious and unexplained presupposition is a weak explanation. It would be *ceteris paribus* better to do without it, which should motivate either a different explanation of metaphysical reference, or a different conception of metaphysics that does not face this problem (deflationism being one, but only one, such conception). To see the structure of the dialectic here, it might be helpful to consider the origin of these problems, Putnam's original permutation argument.

In some presentations, Putnam bases his argument on the Löwenheim–Skolem theorem (LST), which can be stated informally as follows: if a countable first-order theory has an infinite model then it has a model of every infinite cardinality.<sup>91</sup> Let us represent the descriptive content of the terms in our language as a first-order theory; presumably it will be finite, but in any case, not uncountable.<sup>92</sup>

---

<sup>91</sup> Putnam 1980, although in other presentations (e.g., Putnam 1981) the Löwenheim-Skolem theorem drops out. Lewis 1984 (p. 229) argues out that Putnam's argument does not really depend on L-S, but on much simpler considerations about permutations of the domain. For a more precise statement of the theorem, see Putnam 1980.

<sup>92</sup> The descriptive content is a set of reference-fixing truths (see above), e.g., analytic sentences, platitudes, or what have you.

Again, let us follow Sider and assume that the quantificational domain of absolutely fundamental objects is infinite.<sup>93</sup> This means that any theory with an infinite model has a model whose quantificational domain is equinumerous with the set of absolutely structural objects. We can then exploit the bijection between these quantificational domains (guaranteed to exist because they are equinumerous) to generate a model of the theory in the absolutely structural quantificational domain.<sup>94</sup> But then NATO entails that the interpretation with the absolutely structural domain is at least as structural as any other interpretation, including the intended interpretation (reference). So either it is overall more structural than the intended interpretation and we are only talking about the absolutely structural, or it is equally structural as the intended interpretation, in which case there is referential indeterminacy with respect to any term whose meaning is different on the two interpretations. The first disjunct is not a viable option for Sider, for it entails that as long as we speak a (finite or countably infinite) language we refer only to the absolutely structural. The second disjunct is not much better, however, for it entails that there is referential indeterminacy with respect to every term whose meaning differs between the absolutely structural interpretation and the intended interpretation. For reasons that we have already seen, this would generate referential indeterminacy and non-substantive questions in fundamental metaphysics.<sup>95</sup>

It might be objected that this Putnam-inspired argument simply does not apply to Sider. Sider does not think that structuralness is a predicate only of objects (elements of a domain of quantification); it can be applied to items of any syntactic category: predicate-extensions, quantifiers, operators. This means that the model guaranteed to exist by the LST contains absolutely structural objects (e.g., spacetime points), but, Sider can always point out, items of other syntactic categories, e.g., predicate-extensions, will not in general be highly structural (e.g., they will not correspond to

---

<sup>93</sup> On Sider's view it contains spacetime points and sets, both pure and impure (Sider 2011, p. 292).

<sup>94</sup> Thanks to Jack Woods for pushing me on this point.

<sup>95</sup> My argument here bears similarities to that of Williams 2007.



various structural field-properties).<sup>96</sup> So there is no reason to assume that the interpretation with the absolutely structural quantificational domain will be at least as structural in non-quantificational respects as the intended interpretation.

But the burden of explanation here is on Sider. He needs to explain why the interpretation with an absolutely structural domain of quantification (guaranteed to exist by Löwenheim-Skolem) is not at least as overall structural as the intended interpretation (reference). After all, he is the one offering the putative explanation of metaphysical reference. Sider does not offer a worked-out account of comparative structuralness that would allow him to explain why the quantificationally absolutely structural interpretation is less overall structural than the intended interpretation. Consequently, he must simply assume, without explaining why, that it is.<sup>97</sup>

Sider might bite the bullet and embrace what I have at various points assumed he must reject: we only ever refer to the absolutely structural. Firstly, this is so implausible that it only strengthens my argument; if this is Sider's only way out, then he does not have a satisfactory (i.e., not wildly implausible) explanation of the semantic possibility of metaphysics. Secondly, this requires such a revision in our ordinary concept that it is less an answer to our original question (why do terms in metaphysical theories refer?) than it is a simple changing of the topic. Thirdly, it is ineffective against the point (see above) that Sider lacks an explanation why there is not an equally overall absolutely structural interpretation of our language.

I want to acknowledge that Sider does not present this theory of structure as an exhaustive explanation of reference, but as an ingredient in any acceptable theory of reference. Any acceptable theory of reference, Sider argues, needs the notion of structure to constrain the range of possible interpretations. This is fine, but it does not affect my point: Sider's theory of structure, even though it

---

<sup>96</sup> The quantifiers whose domain is absolutely structural is, however, absolutely structural (on Sider's view), because fundamental truths can only quantify over the absolutely structural, and fundamental truths contain only absolutely structural notions; see Sider 2011, 106–9, 292.

<sup>97</sup> The original Putnam permutation argument was that a pure descriptive theory of reference massively under-determines reference. I am making a weaker point against Sider: his structure-augmented descriptivism theory fails to explain why there is *not* massive referential indeterminacy.

may be a *necessary* addition to any other theory of reference, is not *sufficient* to explain metaphysical reference.

The lesson of all of this is that, contra Sider's own description of his 'world view' at the beginning of the book, he does not need to assume merely that the world *has a structure*:

A certain 'knee-jerk realism' is an unargued-for presupposition of this book. Knee-jerk realism is a vague picture rather than a precise thesis. According to the picture, the point of human inquiry--- or a very large chunk of it anyway, a chunk that includes physics---is to conform *itself* to the world, rather than to *make* the world. The world is 'out there,' and our job is to wrap our minds around it. This picture is perhaps my deepest philosophical conviction. I've never questioned it, giving it up would require a reboot too extreme to contemplate; and I have no idea how I'd try to convince someone who didn't share it.<sup>98</sup>

In addition to assuming all of this, Sider must assume additionally *without explanation* that the structure of the world is so constituted as to allow us to refer determinately to the structural absolutely *and not only* to the absolutely structural, or, equivalently, that there is a compact family of overall equally structural interpretations of our language that do not differ with respect to absolutely structural terms in fundamental metaphysics.<sup>99</sup> My point is that there is room to drive a wedge between these: we can admit the world has an objective structure and that reference is essentially structural but question whether the structure of the world and the nature of the reference relation are so aligned as to allow us to do fundamental metaphysics as well as to refer to the merely comparatively structural.

I have argued that various strategies for explaining how our metaphysical terms refer to metaphysical structure in the world are either unsuccessful or ultimately depend on the idea of joint carving, and I have argued that the preeminent theorist of joint carving can explain metaphysical

---

<sup>98</sup> Sider 2011, 18.

<sup>99</sup> I suspect, based on this passage, that Sider would be happy with this description of his project. He is arguing that (on pain of total semantic collapse) our terms must carve at the joints. He does not take himself to be giving an explanation of why. He may even doubt that such an explanation can be given (or that the idea of one is so much as coherent).

reference only by assuming something that stands at least as much in need of explanation.<sup>100</sup> This is a serious explanatory gap in the substantive, non-deflationary view of metaphysics. It should motivate us to look for a meta-metaphysical view that closes the gap: perhaps a non-substantive deflationary view, a substantive view supplemented with enough additional metaphysics to explain reference (which, I have argued, has not *yet* been done) or some alternative not yet considered in the contemporary literature (e.g., Kantian transcendental idealism).<sup>101</sup> In the next section I respond to some skeptical worries about the explanatory gap objection itself.

### §7. Objections and Replies

Before concluding I want to respond to a few pressing objections to how I have set up this (originally Kantian) problem in the first place.

*Not just a problem for metaphysics.* As I noted at the end of Section 2, one likely source of resistance to my argument in this paper is that it proves too much. The same question arises in mathematics and physics: why do the terms in those theories refer? While the dialectical situation is somewhat different than with metaphysics, arguably, many of the same problems canvassed in Section 5 and 6 will arise as well. So this is only a problem for metaphysics insofar as it is a problem for mathematics and the natural sciences.<sup>102</sup> Metaphysics is in no special danger.

But mathematics and natural science are in a dialectically better position because they can appeal to a version of the ‘no miracles’ argument. The enormous success of both sciences over the course of centuries (millennia in the case of mathematics) makes overwhelmingly likely that they are referring to *something* even if we cannot explain *why*.<sup>103</sup> Metaphysics cannot appeal to anything like this history

---

<sup>100</sup> Williams also develops an interpretationist meta-semantics, in which the notion of naturalness plays an important role. I also think Williams’ account fails to explain the semantic possibility of metaphysics, but I do not have the space to argue that here.

<sup>101</sup> For an argument that transcendental idealism can, potentially, do better than inflationism, see Stang Forthcoming-b and Stang 2023.

<sup>102</sup> In fact, a whole set of additional issues arise in the case of mathematics; see Benacerraf 1965 and 1973.

<sup>103</sup> The predictive success of the natural sciences makes it overwhelmingly plausible that they refer to something and state truths about it, and the ineliminability of mathematics from the natural sciences makes it overwhelmingly likely that mathematics also refers to and states truths. I am not trying to take a stand here on whether mathematics is genuinely ineliminable from natural science, but only to

of successes, so it stands in a different dialectical position with respect to its (alleged) failure to explain the reference of its own terms. Metaphysics stands under the suspicion of being semantically defective, of involving terms that do not refer. So even if the same meta-semantic issues can be raised about the reference of mathematical or natural scientific terms, the lack of a satisfying explanation is more dialectically problematic for the metaphysician than it is for the mathematician or the natural scientist. We do not need want to know merely *why* metaphysical terms refer, but *whether* they do at all, or whether metaphysics is semantically defective.

*Ethics and metaphysics.* Likewise, analogous problems can also be raised about the reference of moral terms. That this should be so is not surprising; in addition to his critique of theoretical reason (his investigation into the possibility of theoretical metaphysics), Kant also offers a critique of *practical* reason, i.e., an investigation into the question, how is morality (in the form of a categorical imperative) possible?<sup>104</sup> But, tellingly, contemporary meta-ethicists have not responded to questions about the reference of ethical terms by simply pleading that these are problems that face other parts of philosophy. Instead, they have developed elaborate and sophisticated answers. Some have offered substantive accounts of how we refer to moral properties, whether these are understood naturalistically (e.g., Brink 1986) or as non-natural (e.g., Shafer-Landau 2003). Others have argued that moral claims purport to refer to non-natural properties, but, because there are no such properties, they suffer from systematic reference-failure (e.g., Mackie 1977). Yet others have argued that moral discourse is not cognitive, and therefore not in the business of referring in the first place (e.g., Stevenson 1963), or offered quasi-realist explanations of the reference of moral terms that apply even if moral discourse is not cognitive (e.g., Blackburn 1993). It seems to me that (inflationary) metaphysicians should do the same: not merely plead that other areas of philosophy face analogous semantic problems but attempt resolutely to face and solve them.

---

point out that this a strong prima facie defense of the semantic possibility of mathematics and natural science to which metaphysics cannot appeal.

<sup>104</sup> This critique of practical reason is also a critique of the possibility of practical metaphysics (the metaphysics of morals), according to Kant's conception of metaphysics (A841/B869).

*Meta-semantics and metaphysics.* Some readers will reply that my problem is not a problem in *metaphysics*, but in *meta-semantics*, to which I have two responses. First, I think of meta-semantics as part of metaphysics, the part that studies the grounds of semantic facts.<sup>105</sup> So a problem in meta-semantics is a problem in metaphysics. Second, philosophical questions cannot be answered simply by changing one's hat. If, as I have argued, it is correct that there is an unsolved problem about the semantic possibility of (inflationary) metaphysics, we should seek a solution to it, rather than passing the buck to someone down the hall ('I'm a metaphysician, that's not my business'). Furthermore, if you agree that meta-semantics is part of metaphysics, then explaining the semantic possibility of natural science is also a task for metaphysics. So any putative metaphysical explanation of the semantic possibility of natural science will naturally raise the question of its own semantic possibility (i.e., why the terms in that very theory refer). Therefore, the observation, discussed above, that the same semantic problem arises for reference in natural science further supports my conception of the dialectic: metaphysics is going to have to explain its own semantic possibility in order to explain satisfactorily the reference of natural scientific terms (for otherwise the very same questions will arise about the very terms in which that explanation is couched), so it cannot put off the question.

*Semantics.* Nor, as some readers might worry, do my arguments undermine the semantic status of various sciences that use the notion of reference, e.g., linguistic semantics. To adopt a toy model of linguistics for the moment, we can think of the linguist as assuming certain reference-facts she takes to be primitive (e.g., the reference of certain lexically primitive items) and using those to explain further reference-facts (e.g., through a compositional theory of the reference of lexically complex items). The meta-semanticist comes along and gives an explanation of the reference-facts taken to be primitive by the linguist, and part of that explanation may be that reference is structural. But then my questions in Section Six arise. These are problems for the meta-semantic-cum-metaphysical

---

<sup>105</sup> If you worry that this turns semantics into metaphysics, see the next paragraph.

explanation of the semantic facts that the linguist uses. They are not problems in linguistics, though they may be problems in the metaphysical foundations of linguistics.<sup>106</sup>

### §8. Conclusion

This brings us back to where this essay began, Kant's critique of metaphysics. In the Preface to the *CPR*, Kant urges reason to 'take on anew the most difficult of all tasks, namely, that of self-knowledge, to institute a court of justice, by which reason may secure its rightful claims while dismissing all its groundless pretensions, and this not merely by decrees but according to its own eternal and unchangeable laws, and this court is none other than the *critique of pure reason* itself' (Axi). Reason undertakes this self-examination in order to answer the question of whether metaphysics is (semantically, epistemically, logically, etc.) possible. To continue Kant's judicial metaphor, reason is on trial because its loftiest creation, metaphysics, has not yet entered upon the 'secure path of a science' (Bi). That metaphysics is not yet a science is evident, Kant thinks, from the fact that metaphysics is a 'battlefield of endless controversies' (Aviii), in which 'it proves impossible for the different co-workers to achieve unanimity as to the way in which they should pursue their common aim' (Bvii), where no results are established without immediately being contradicted by other practitioners: 'on this battlefield no combatant has ever gained the least bit of ground, nor has any been able to base any lasting possession on his victory' (Bxv). If this resonates with the reader, as it does with the author, as a description of the contemporary state of metaphysics, then metaphysics is 'in the dock' while the established sciences (Kant mentions logic, mathematics, and physics) are not. Metaphysics is under a cloud of suspicion that the other sciences are not, and its failure to explain why its terms refer threatens to make reasonable a verdict that would be unjustified in the case of the

---

<sup>106</sup> Note that this means that if metaphysics is semantically impossible (if our metaphysical terms do not carve the world at its ontological joints) this will entail, at the worst, that we cannot explain why our physical terms refer. It will not entail that they fail to refer, or that there is massive semantic indeterminacy. It will not mean what Sider (following Fodor) calls 'the end of the world.'

other sciences: that it is semantically defective, and this is why it has never entered upon the secure path of a science.<sup>107</sup>

In Section Two I pointed out several reason for thinking that Kant's critique of metaphysics may have been rendered obsolete by later developments:

1. It rests on the analytic-synthetic distinction
2. It assumes that metaphysics must be *a priori*
3. It is about 'big M' Metaphysics
4. It assumes metaphysics must be knowledge
5. It depends on an 'epistemology first' methodology
6. It depends upon the notion of 'intuition'
7. It assumes that metaphysics is about the supersensible

But we have seen that Kant's central question about metaphysics ('the key to the whole secret of metaphysics'), the problem of the objective validity of the categories, which I have interpreted as the problem of why the basic concepts of metaphysics refer, depends on none of these. So, to the question in the title of this essay, I answer: no!

---

<sup>107</sup> 'As far as metaphysics is concerned, however, its poor progress up to now, and the fact that of no metaphysics thus far expounded can it even be said that, as far as its essential end is concerned, it even really exists, leaves everyone with ground to doubt its possibility' (B21).

## Works Cited

- Adams, F. & Aizawa, K. 2017. Causal Theories of Mental Content. *The Stanford Encyclopedia of Philosophy*. E. Zalta (Ed.). URL = [<https://plato.stanford.edu/archives/sum2017/entries/content-causal/>](https://plato.stanford.edu/archives/sum2017/entries/content-causal/).
- Bacon, A. 2018. The Broadest Necessity. *Journal of Philosophical Logic* 47 (5):733-783.
- Benacerraf, P. 1965. What numbers could not be. *Philosophical Review* 74 (1):47-73.
- Benacerraf, P. 1973. Mathematical truth. *Journal of Philosophy* 70 (19):661-679.
- Blackburn, S. 1993. *Essays in Quasi-Realism*, New York: Oxford University Press.
- Blatti, S. & Lapointe, S. (Eds.). 2016. *Ontology After Carnap*. Oxford: Oxford University Press.
- Brink, D. 1986. Externalist Moral Realism. *Southern Journal of Philosophy* (Supplement), 24: 23–42.
- Burgess, A. & Sherman, B. (Eds.). 2014. *Meta-Semantics: New Essays on the Foundations of Meaning*. Oxford: Oxford University Press.
- Campbell, K. *Abstract Particulars*. Oxford: Blackwell.
- Carnap, R. 1950. Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie* 4 (11), 20-40.
- Chalmers, D., Manley, D., & Wasserman, R. (Eds.). 2009. *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.
- Coffa, A. 1991. *The Semantic Tradition from Kant to Carnap: To the Vienna Station*. L. Wessels (Ed.). Cambridge: Cambridge University Press.
- Della Rocca, M. 2010. PSR. *Philosophers' Imprint* 10.
- Devitt, M. 1981. *Designation*. New York: Columbia University Press.
- Eklund, M. 2009. Carnap and Ontological Pluralism. In Chalmers et. al. 2009, 130–156.
- Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: Harvard University Press.
- Friedman, M. 1992. *Kant and the Exact Sciences*. Cambridge, MA: Harvard University Press.
- Hanna, R. 2006. *Kant, Science, and Human Nature*. Oxford: Oxford University Press.
- Hawthorne, J. 2006. *Metaphysical Essays*. Oxford: Oxford University Press.



- Hirsch, E. 2009. Ontology and Alternative Languages. In Chalmers et. al. 2009, 231–259.
- 2010. *Quantifier Variance and Realism: Essays in Ontology*. Oxford: Oxford University Press.
- Hofweber, T. 2009. Ambitious, Yet Modest Metaphysics. In Chalmers et. al., 260–289.
- 2016. *Ontology and the Ambitions of Metaphysics*. Oxford: Oxford University Press.
- Hogan, D. 2009. Three kinds of rationalism and the non-spatiality of things in themselves. *Journal of the History of Philosophy* 47 (3), 355–382.
- Jackson, F. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Kant, I. 1998. *Critique of Pure Reason*. P. Guyer & A. Wood (Eds. and Trans.). Cambridge: Cambridge University Press.
- 1999. *Correspondence*. A. Zweig (Ed. and Trans.). Cambridge: Cambridge University Press.
- 2002. *Theoretical Philosophy after 1781*. H. Allison & P. Heath (Eds. and Trans.). Cambridge: Cambridge University Press.
- 2004. *Der Streit mit Johann August Eberhard*. Ed. M. Lauschke & M. Zahn. Frankfurt: Meiner.
- Kim, J. 1973. Causation, Nomic Subsumption, and the Concept of an Event. *The Journal of Philosophy* 70(8): 217–236.
- Kriegel, U. 2013. The Epistemological Challenge of Revisionary Metaphysics. *Philosophers' Imprint* 13(12), 1–30.
- Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Ladyman, J. & Ross, D. 2007. *Everything Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Lewis, D. 1973. Causation. *Journal of Philosophy* 70: 556–67.
- 1983. New work for a theory of universals. *Australasian Journal of Philosophy* 61: 343–77.
- 1984. Putnam's Paradox. *Australasian Journal of Philosophy* 62(3): 221–236.
- 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
- 1991. *Parts of Classes*. Oxford: Blackwell.

- Lowe, E.J. 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford: Oxford University Press.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.
- Maudlin, T. 2007. *The Metaphysics within Physics*. Oxford: Oxford University Press.
- McDaniel, K. 2007. Extended Simples. *Philosophical Studies* 133:131–141.
- McGee, V. 2005. Two Conceptions of Truth? *Philosophical Studies* 124: 71–104.
- Pruss, A. 2009. *The Principle of Sufficient Reason: A Reassessment*. Cambridge: Cambridge University Press.
- Putnam, H. 1975a. *Philosophical Papers, Vol. 1: Mathematics, Matter, and Method*. Cambridge: Cambridge University Press, 131–193.
- 1975b. The Meaning of ‘Meaning.’ In K. Gunderson (Ed.). *Language, Mind, and Knowledge*. Minnesota Studies in the Philosophy of Science VII. Minneapolis, MN: University of Minnesota Press.
- 1977. Realism and Reason. *Proceedings of the American Philosophical Association* 50, 483–498. Repr.
- Putnam, H. 1978. *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.
- 1980. Models and Reality. *Journal of Symbolic Logic* 45, 464–482.
- 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Quine, W.V.O. 1951. Two Dogmas of Empiricism. *Philosophical Review*, 60: 20–43.
- Rosefeldt, T. 2013. Subject-Dependence and Trendelenburg’s Gap. In M. Ruffing, C. La Rocca, A. Ferrarin & S. Bacin (eds.). *Kant Und Die Philosophie in Weltbürgerlicher Absicht: Akten des Xi. Kant-Kongresses 2010*. Berlin, De Gruyter. 755–764.
- Russell, G. 2008. *Truth in Virtue of Meaning: A Defense of the Analytic/Synthetic Distinction*. Oxford: Oxford University Press.
- Schaffer, J. 2016. Grounding in the image of causation. *Philosophical Studies* 173 (1):49-100.
- Shafer-Landau, R. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.

- Sider, T. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- 2009. Ontological Realism. In Chalmers, et. al., 384–423.
- 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
- Simchen, O. 2017. *Semantics, Meta-Semantics, and Aboutness*. Oxford: Oxford University Press.
- Stang, N. 2018. How is Metaphysics Possible? Kant's Great Question and His Great Answer. In Stephen Hetherington (ed.), *What Makes a Great Philosopher Great? Thirteen Arguments for Twelve Philosophers*. Routledge.
- Stang, N. Forthcoming-a. IX—How Is Metaphysics Possible? *Proceedings of the Aristotelian Society*.
- Stang, N. Forthcoming-b. Why Should Metaphysics be Systematic? Contemporary Answers and Kant's. In *Systematic Metaphysics: Contemporary and Historical Perspectives*. Eds. Aaron Segal & Nicholas Stang. Oxford University Press.
- Stang, N. 2023. Metaphysics on the Model of Natural Science? A Kantian Critique of Abductivism.' In *Metaphysics as a Science in Classical German Philosophy*. Ed. Robb Dunphy & Toby Lovatt. (London: Routledge), pp. 339–366.
- Stevenson, C.L. 1963. *Facts and Values: Studies in Ethical Analysis*. New Haven: Yale University Press.
- Strawson, P.F. 1966. *Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London: Methuen.
- Thomasson, A. 2009. Answerable and Unanswerable Questions. In Chalmers, et. al., 444–471.
- 2014. *Ontology Made Easy*. New York: Oxford University Press.
- Van Fraassen, B. 2002. *The Empirical Stance*. New Haven: Yale University Press.
- Weatherson, B. 2003. The Role of Naturalness in Lewis's Theory of Meaning. *Journal for the History of Analytical Philosophy* 1(10): 1–19.
- Westphal, K. 1989. *Hegel's Epistemological Realism: A Study of the Aim and Method of Hegel's Phenomenology of Spirit*. Dordrecht: Kluwer.

- Williams, J. R. G. 2005. *The Inscrutability of Reference*. Ph.D. thesis, University of St Andrews. Available online at <https://webpace.utexas.edu/deverj/personal/test/inscrutabilityofreference.pdf>.
- 2007. Eligibility and Inscrutability. *Philosophical Review*, 116: 361–99.
- 2019. *The Metaphysics of Representation*. Oxford: Oxford University Press.
- Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
- 2003. Everything. *Philosophical Perspectives* 17(1): 415–465.
- 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.
- 2013. Logic, Metalogic, and Neutrality. *Erkenntnis* 79(supp. 2): 211–231.
- 2016. Modal science. *Canadian Journal of Philosophy* 46 (4-5): 453-492.
- Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. C. K. Ogden (Trans.). London: Routledge and Kegan Paul Ltd. English Translation of 1921. *Logisch-Philosophische Abhandlung. Annalen der Naturphilosophie*, Ostwald, 1921.
- Yablo, S. 2009. Must Existence Questions Have Answers? In Chalmers et. al. (2009), 507–525.