

## Is Kant's Critique of Metaphysics Obsolete?

Nick Stang, University of Toronto  
[nick.stang@gmail.com](mailto:nick.stang@gmail.com)

**Draft—do not circulate or cite without permission!**

### Abstract

I raise a problem about the possibility of metaphysics originally raised by Kant: what explains the fact that the terms in our metaphysical theories (e.g. “property”) refer to entities and structures (e.g. properties) 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 by Leibniz and Wolff.<sup>2</sup> Kant’s critique had allegedly been rendered obsolete, according to Kant’s

---

<sup>1</sup> *Über eine Entdeckung nach der alle neue Kritik der reinen Vernunft durch eine ältere*

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

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.<sup>3</sup>

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>4</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>5</sup> This has given rise to a feeling among some philosophers that analytic metaphysicians are, mistakenly, returning to a kind of

---

<sup>3</sup> By 'analytic metaphysics' I mean, roughly, metaphysics as it has been practised in analytic philosophy since the pioneering work of David Lewis.

<sup>4</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 what 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>5</sup> See Pruss 2009 & Della Rocca 2010.

metaphysics that Kant showed to be impossible.<sup>6</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 previous generations took it to be.<sup>7</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 untenable, or that metaphysics cannot obtain a certain kind of epistemic status (a priori knowledge, certainty, the "secure path of a science," etc.).<sup>8</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

---

<sup>6</sup> E.g. van Fraassen 2002, 2–4.

<sup>7</sup> See Williamson 2007, 19; Lowe 1998, 1–8.

<sup>8</sup> Some of these points are raised by Lowe 1998.

analytic/synthetic distinction and have thus been skeptical of any philosophical project formulated in terms of one half of it.<sup>9</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>10</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>11</sup> So if Kant's arguments about metaphysics rest on the assumption that metaphysics must be strictly a priori, they seem outdated.<sup>12</sup>

---

<sup>9</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). The important point, for our purposes, is that few agree that metaphysics as such must be either analytic (Leibniz) or synthetic (Kant), so the fate of the analytic/synthetic distinction might be thought to be irrelevant, pace Kant, to the possibility of metaphysics.

<sup>10</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>11</sup> E.g. Maudlin 2007.

<sup>12</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>13</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 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

---

<sup>13</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).

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>14</sup>

5. *Apodictic certainty.* Related to the issue of the a priori, Kant conceives of metaphysics as aiming at apodictic certainty. For this reason, he claims that metaphysics cannot rest on hypotheses or judgments about what is probably the case.<sup>15</sup> But few contemporary 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>16</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 astro-physics, 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

---

<sup>14</sup> See Lewis 1986, 108–115; Sider 2011.

<sup>15</sup> Axv, A774/B802.

<sup>16</sup> Kant 2002, 353.

absent from the contemporary scene.<sup>17</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" (*Anschaung*) and his controversial views about the dependence of our knowledge on intuition. Since the first is tied up with Kant's arguments in the Transcendental Aesthetic, which are notoriously controversial, and the 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

---

<sup>17</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.

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.

---

<sup>18</sup> Indeed, one can think of the whole 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> I am currently writing a book on the topic.

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 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. Several 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 are forthcoming. The closest thing to a response is the notion of

---

<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; much of my Kant interpretation that is relevant here is unpublished, but see Stang 2016, 158–170, and “Kant’s Schematism of the Categories: An Interpretation and Defense” (unpublished MS) for some relevant elements.

“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 to *either* 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.

## §2. The Reference of Metaphysical Concepts

Allow me one Kant quote:

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>22</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>23</sup>

Kant’s question then becomes: what explains (“on what ground rests”) the

---

<sup>22</sup> Kant 1999, 133. Translation slightly modified by author.

<sup>23</sup> I won’t defend the interpretive point but I do so in forthcoming work.

reference of our mental representations to objects?<sup>24</sup> Kant goes on in, the rest of the letter, 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 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>25</sup>

---

<sup>24</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.

<sup>25</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). See Kriegel (2013) for a contemporary version of the Kantian challenge to the epistemic possibility of (revisionary) metaphysics.

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, its terms refer to entities and structures in the world.<sup>26</sup> What explains that fact?

By a “referential semantics” I mean one that contains familiar Tarski-style recursive truth-conditions for atomic sentences, quantification, quantifiers, logical connectives, etc. We have a domain  $D$  and a function  $R$  (the *reference* function) that maps 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> 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 “meaning” of the universal quantifier  $\forall$ . We can ask, for instance, why  $\forall$  has this meaning, rather than another. 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>27</sup>

In this way, questions of reference can also be applied to terms in metaphysical theories (henceforth, metaphysical term) 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 as the modal primitivist includes in her semantic theory some clause of the form:

---

<sup>27</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.

(1) “Possibly  $p$ ” is true if and only if it is possible that  $p$ .<sup>28</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 referring can be generalized to metaphysical terms that are neither singular terms (names, definite descriptions, etc.) nor predicates nor quantifiers.<sup>29</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>30</sup>

The Kantian objection can be concisely stated:

---

<sup>28</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?

<sup>29</sup> Sider 2011, 100–123.

<sup>30</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.

- (P1) If any metaphysical theory is true then its terms refer, either to entities in the world or to worldly structure.
- (P2) There is no explanation of why terms in metaphysical theories refer, either to entities in the world or to worldly structure.
- (C) ∴. Either no metaphysical theory is true or there is an unexplainable fact, the fact of metaphysical term reference.

In the next section I will argue that (P2) does not apply to all metaphysicians; theorists who adopt a deflationary meta-metaphysics (in a sense to be specified) can easily explain why metaphysical concepts refer.<sup>31</sup> The conclusion of the argument must be restricted accordingly: either no *substantive* metaphysical theory is true etc. The substantive metaphysician must either give up claiming her theory to be true or must accept an unexplained fact, the fact of metaphysical term reference. My method will be to argue by elimination: none of the extant theories of reference explain non-deflationary metaphysical term reference. 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. Whether this should be accomplished by offering a new theory of reference within substantive meta-metaphysics, or embracing some kind of deflationary meta-metaphysics, is an issue I will postpone for future work.<sup>32</sup>

---

<sup>31</sup> Kant also thought that P2 was not true of his his own “critical” metaphysics, but that lies outside the scope of this paper. See my “Kant’s Schematism of the Categories” (unpublished MS).

<sup>32</sup> It should also be noted that (P1) could be rejected by a metaphysician who does not accept a referential semantics for her metaphysical theory at all. For instance, a metaphysician could adopt an expressivist semantics for metaphysics in which the basic

Finally, to anticipate a worry I imagine many readers will have: while any discourse for which we adopt a referential semantics incurs a similar explanatory burden, and it is notoriously hard to give a unified explanatory theory of reference, I will argue that explaining the reference of metaphysical terms is especially problematic for two reasons. First of all, many theories of reference were developed to deal with terms of a certain sort (e.g. names) and are simply not applicable to metaphysical terms. Secondly, metaphysics is especially problematic because the best explanations of the reference of terms in other discourses are metaphysical explanations (e.g. explanations in terms of joint-carving). Thus metaphysics serves as an explanatory “backstop” to the referential semantics of other theories (e.g. physics). Metaphysics, I will argue, has no further backstop to appeal to and cannot backstop itself.

### **§3. Specifying the Target of Kant’s Critique:**

#### **Substantive vs. Deflationary Metaphysics**

---

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 concoct 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. 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.

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 (they have enough names, predicates, etc. to give a “complete” description) languages are metaphysically “on a par,” and on the other hand, views according to which some such theories do a better job at “carving” the world “at its joints” (Sider 2011). Alternatively, one can distinguish between views on which metaphysical 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 metaphysical 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>33</sup>

My way of dividing up the logical space is slightly different. I distinguish between *substantive* views, according to which reference cannot be explained by reflection on semantic rules of our language, from *deflationary* views, according to which the reference of our metaphysical terms can be trivially explained.<sup>34</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

---

<sup>33</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>34</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.

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>35</sup> If meta-semantics is part of metaphysics then my meta-semantic objection to 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>36, 37</sup>

But why does the alleged lack of an explanation of metaphysical term reference pose a problem for the substantivist? It is not bad merely because unexplained facts are bad (though they are), but because explained facts about the reference of words in our language are especially bad. This is a blunt appeal to intuition: among the fundamental furniture of the universe there are no facts of the form “term *x* in metaphysical theory *T*

---

<sup>35</sup> C.f. “metasemantics is the business of providing metaphysical explanations of semantic facts” (Burgess & Sherman 2014, 22).

<sup>36</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>37</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.

refers to *y*.” Additionally, I will argue in Section Six that a prominent proponent of the substantivism, Ted Sider, cannot, even within his own picture, accept fundamental reference facts. The explanatory gap would be, I take it, a real problem for the substantivist.

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>38</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.

Although Thomasson does not, to my knowledge, explicitly address the question of why metaphysical terms refer,<sup>39</sup> it is not hard to construct one 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

---

<sup>38</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>39</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.

empirical facts. For instance, the term “table” is governed by this semantic rule, in which it is used rather than mentioned:

(S<sub>O</sub>) If there are particles arranged table-wise then there is a table.

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>40</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<sub>O</sub> is the object-language correlate (i.e. it *uses* “table”) of the following meta-language rule (i.e. it *mentions* “table”):

(S<sub>M</sub>) If there are particles arranged table-wise then “table” refers.

The explanation of why “table” refers works along now familiar lines: S<sub>M</sub> 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<sub>M</sub>) nor of why “there are tables” is true (the truth of the right-hand side of S<sub>O</sub>). That “table” refers is a logical consequence of analytic sentence (S<sub>M</sub>) and an undisputed empirical truth that does not implicitly or explicitly quantify over tables (according to Thomasson).<sup>41</sup> On

---

<sup>40</sup> Thomasson follows Kant and Frege in identifying *what exists* with *what there is* (Thomasson 2014, 85–86).

<sup>41</sup> Understanding 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

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 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.

---

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."

- (5) Colors are properties.
- (6) ∴ There is a property.
- (7) There is a property iff “property” refers.
- (8) ∴ “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>42</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, 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 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

---

<sup>42</sup> It is empirical because it entails that Beyoncé’s dress exists, which, presumably, is not something one can know a priori.

face the explanatory gap that Kant pointed out.<sup>43</sup> This means that the Kantian objection must be appropriately restricted in its scope: it applies only to the substantive metaphysical view. In the remainder of this paper, “metaphysics” and “metaphysicians” should be understood as appropriately restricted.

#### **§4. Dismissing the Explanatory Demand**

Before continuing I to amend slightly (P2) 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>44</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 terms using the reference of the terms from which it is defined. This means that we should reformulate P2 as:

---

<sup>43</sup> Which does not mean they are unproblematic by Kantian lights. I criticize neo-Carnapian views in “Old Problems for New Carnapians” (unpublished MS).

<sup>44</sup> See McDaniel 2007.

(P2\*) Within a substantive meta-metaphysics, there is no explanation of why primitive metaphysical terms refer, either to objects in the world or to worldly structure.<sup>45</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  $(a, \{a\})$ ,<sup>46</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 question and claims can be formulated in the

---

<sup>45</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>46</sup> Lewis 1991 attempts to build all of set theory out of mereology plus the primitive relation of an individual to its singleton-set.

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>47</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 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 own 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). Even if many metaphysical terms (e.g. “property”) are used in other fields, the question must be faced: why do they refer? As I will argue in this section and the next, there is no easy answer to that question, for such metaphysical terms, even if they also figure in the discourses of other fields.

2. *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

---

<sup>47</sup> For contrasting views, see Wittgenstein 1922 and Williamson 2013.

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 is that Fregean concepts have their referents essentially; 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.). 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.

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 property.<sup>48</sup> But I do not think the conjunction of those facts explains the fact that  $r$ . In general, it is intuitive that, merely having explained  $p$  and  $q$ , where  $p \ \& \ q$  entails  $r$ , we have not thereby explained  $r$ . For instance, intuitively, having explained the fact that Stephen King was able to write 2 novellas and 4 short stories in 1982 ( $p$ ) and the fact that he was able to write a novel in 1982 ( $q$ ) we have not thereby explained something entailed by the conjunction of these facts, namely, that Stephen King was able to write 1 novel, 2 novellas, *and* 4 short stories in a single year! The explanation of  $p$  might explain how King found the time to write 2 novellas and 4 short stories, while  $q$  would explain how he found the time to write a novel; but conjoining those explanations puts us in no better position to understand how he wrote

---

<sup>48</sup>  $r = \exists x(\text{“substance” means } x \ \& \ x \text{ refers to substance}).$

everything he did in 1982. In general, then, separate explanations of conjuncts do not add up to explanations of conjunctive facts.

Moreover, in the case at hand, separate explanations of the conjuncts leave unexplained their connection: the psychological explanation of  $p$  (by hypothesis) 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 *pro tanto* wrong) and a purely normative-ethical explanation of why those beliefs are true (e.g. our belief that incest is wrong from some more basic moral principle), but we would not thereby have explained why we hold true ethical beliefs. 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.

3. *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 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 paper, 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>49</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 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**

---

<sup>49</sup> Originally adumbrated by Putnam 1975*a*, 73.

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>50</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>51</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 casual 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

---

<sup>50</sup> E.g. Kripke 1980, Devitt 1981.

<sup>51</sup> See the summary of these problems in Adams & Aizawa 2017.

views of causation, properties play an important role in causal relations<sup>52</sup>; indeed on some views, they are the principal causal relata.<sup>53</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

---

<sup>52</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>53</sup> Campbell 1990.

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>54</sup>

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 under-statement: 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

---

<sup>54</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 reference is a theory specifically about *causal* determination, not determination in general: the referent of a term is what *causally* determines its tokenings.

strategy for defending metaphysics. No wonder then, that no extant causal theory of reference has attempted it.<sup>55</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 follows from 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” 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

---

<sup>55</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).

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>56</sup> Global

---

<sup>56</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. Sider 2011, 28–32, shows that the

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 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 (truth-tracking) mental state, the most basic way in which our minds are related to reality.<sup>57</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

---

argument does not depend on the Löwenheim-Skolem theorem (as Putnam claims), but instead on very simple considerations of elementary model theory.

<sup>57</sup> Williamson 2007, 269; for a more comprehensive defense of the “knowledge first” approach, see Williamson 2000.

dialectically well placed to answer the Kantian question about metaphysical term reference. Williamson's view applies to our language *as a whole*.<sup>58</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 answer may be 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 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

---

<sup>58</sup> He does so in order to avoid otherwise intuitively compelling counter-examples; see Williamson 2007, 267.

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>59</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>60</sup> My account of joint carving 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

---

<sup>59</sup> I address this issue in work currently in preparation.

<sup>60</sup> Williams 2005 contains a “naturalness”-based theory of reference, but I will focus on Sider 2011 because it is better known. Williams 2007 raises some objections to the details of Sider’s view; my objections are pitched at a greater level of generality. Weatherson 2003 raises some problems for the standard reading of Lewis. Cf. Hawthorne 2006, 58–61, for further discussion of naturalness and reference magnetism.

the joints (the original source of Plato's metaphor<sup>61</sup>) or fail to do so (e.g. by quartering it), we can "carve" the 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 plays an important role in Sider's theory of reference.<sup>62</sup> The descriptive content of a term (call this its *meaning*) determines a range of "reference candidates," those that make true the descriptive content of the term. 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>63</sup> The actual referent of the term is the reference-candidate that is maximally joint carving (in the case of "water," H<sub>2</sub>O). If there is no such unique candidate (if multiple candidates are equally joint carving) 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 "nonsubstantive." For instance, whether bachelor refers to all unmarried males (including the pope) or only to unmarried males eligible for marriage (thus not the pope)

---

<sup>61</sup> *Phaedrus*, 265e-266a.

<sup>62</sup> Sider 2011's preferred term is "structure" because he does not want to be ontologically committed to joints. However, I will use "joint-carving" because I find it expresses the same idea more intuitively and it can be read so as not to incur ontological commitment to joints.

<sup>63</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.)

is a nonsubstantive question because each of the reference-candidates for “bachelor” (the property *unmarried male*, the property *unmarried male eligible for marriage*) is equally joint carving (that is, not especially).<sup>64</sup>

Can Sider’s theory of joint carving explain how metaphysical terms refer? Take some metaphysical term like “particular” and let us assume it carves perfectly at the joints. (I am not sure Sider thinks “particular” is joint carving, but this does not matter to the example.) “Particular” refers because there is a unique reference-candidate that satisfies the descriptive meaning of the term (particulars instantiate properties but are not themselves instantiated) and is maximally joint carving (let us assume in this case), the set of all particulars (or the property of being a particular—this difference will not matter). Now why does the reference relation hold between the term “particular” and the set of particulars? Because the set of particulars is (i) a reference-candidate for “particular” (ii) carves at the joints.<sup>65</sup>

The reference relation is one among (provably) many assignments of terms in our language to objects and structures in the world. Why does it map terms in our language to

---

<sup>64</sup> Sider 2011’s preferred term is “structure” because he does not want to be ontologically committed to joints. However, I will use “joint-carving” because I find it expresses the same idea more intuitively and it can be read so as not to incur ontological commitment to joints.

<sup>65</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 reference of our terms are “natural” (Lewis) or “structural” (Sider), so I will largely dispense with this metaphor in what follows.

objects and structures that carve the world at its joints? (In Sider's terms, why are the referents of our terms "structural"?) As I see it, Sider has five responses open to him.

*A. Reference as reference magnet.* Call the semantic relation (a relation between terms in our language and entities and structure in the world) that Sider thinks is reference R. Let R\* be a semantic relation that assigns non-joint-carving semantic values to the terms in our language while preserving truth (i.e. one of Putnam's permuted assignments). Now let us ask: why does the term "reference" refer to R rather than R\*?

Let us represent this putative fact as Refers ("reference," R).<sup>66</sup> Sider's first strategy for explaining why reference is a joint-carving relation works by pointing out that the semantic relation R that maps our words to joint-carving entities and structures is itself a more joint-carving relation than R\*. So, in particular, this means that R maps "reference" to itself, i.e. R ("reference," R), rather than R\*. But this fact is the fact we want to explain, namely that "R" refers to reference, *only if* Refers = R (i.e. if "refers" refers to R). In other words, in order to explain the fact that "R" refers to reference, Sider has to assume that R *is* reference (or that "refers" refers to R). So Sider's explanation is ultimately circular, that is, no explanation at all. We start off wanting to know why "reference" refers to a joint-carving relation (why Refers("reference," R) rather than Refers("reference," R\*)) and we are told this is because R is a joint-carving relation. But what has R to do with reference? Why does the fact that R maps "reference" to R itself have to do with the putative fact that "reference" refers to R? This does issue in an

---

<sup>66</sup> To be read " 'reference' bears the reference relation to R" or, more simply, " 'reference' refers to R."

explanation if we assume that Refers (“reference,” R) but that is precisely what we wanted an explanation of in the first place. So Sider’s first strategy fails as an *explanation* of why reference carves at the joints.<sup>67</sup>

*B. Semantic Explanation.* Facts about what our words refer to figure in explanations. For instance, to use an example Sider borrows from Van McGee: why did the teacher write the sentence “Maiasurs were highly social animals that traveled in herds of as many as 10,000”?<sup>68</sup> Part of the explanation is that “Maiasurs” refer to maiasurs; the rest being some facts about the teacher’s intentions, beliefs, etc. Explanations, on Sider’s view, must carve at the joints, or, at least, an explanation is better the more it carves at the joints. Since “reference” is part of the explanation of semantic explanations, “reference” must carve at the joints. But note that this is at most a reason to think *that* reference carves at the joints, not an explanation of *why* this is the case.

But it fails even as an argument that reference is joint-carving. The conclusion only follows if we assume that what reference itself explains is relatively joint-carving. In the “maiasaur” argument, the explananda are facts about the author’s semantic intentions, etc. In general, reference facts explain broadly semantic facts about what we are talking

---

<sup>67</sup> Throughout, when I say of a semantic relation (R, R\*, etc.) that it is “joint carving” I mean that it maps terms in our language to relatively natural or “structural” entities and structures in the world (to use Sider’s terms). The naturalness/structuralness/joint-carving-ness of reference (or R, R\*, etc.) is the naturalness/structuralness/joint-carving-ness of the *values* it assigns to terms in our language, not of the relation itself.

<sup>68</sup> McGee 2005, section 4. Quoted by Sider 2011, 28.

about, which objects are relevant to the evaluation of the truth of our beliefs, why certain actions were rational (e.g. writing the “maiasaurs” sentence, etc.). If the semantic facts that reference is used to explain are themselves highly non-joint carving (perhaps they are as gerrymandered as one of Putnam’s permuted models) then there is no reason, even within Sider’s system, to think that reference carves at the joints. To put the same point another way, if semantic notions like “aboutness” are not joint-carving then there is no substantive question about what they refer to; there is a plurality of equally joint-carving candidates for our semantic notions. If we then want to explain those semantic notions using the core semantic concept of reference, there is no reason to think there will be a unique, much less a unique and *joint-carving*, relation fitting the descriptive role of “reference” that explains these non-substantive semantic facts. So in order to explain why reference is a joint carving relation (e.g. why “reference” refers to R rather than R\*) Sider must assume that some *other* term in a family of semantic facts (e.g. aboutness, truth-conditions, etc.) carves at the joints. But that is merely to push the explanatory requirement back one step, not to answer satisfactorily the question, why is reference joint carving?

*C. Stipulation.* Alternately, Sider can stipulate that reference carves at the joints by building it into the descriptive meaning of the term: let “reference” be the joint carving relation such that (insert a description of the theoretical role of reference within semantics). We can add this stipulation, if we wish, to Sider’s meta-semantic theory (this theory of what semantic notions like “reference” refer to). But this will only succeed if “joint carving” itself carves at the joints, i.e. if “joint carving” refers to joint carving (JC,

for short) rather than some other structural property like joint carving\* (JC\*), the property a term has if it refers to entities that are non-structural. But this means that Sider's stipulation strategy only succeeds if at least one term in our language carves at the joints, either "reference" or "joint carving" itself. A community whose language did not carve at the joints at all could not "bootstrap" themselves into joint-carving by stipulating that the terms in their language <sup>^</sup>carve at the joints<sup>^</sup>, where <sup>^</sup>carves at the joints<sup>^</sup> is an expression in their *ex hypothesi* non-joint-carving language. So the explanatory power of this stipulative strategy is parasitic on the assumption that (at least some, e.g. "joint carving" itself) terms in our language carve at the joints. It is thus not capable of itself explaining why that assumption is true.

*D. Definitional.* Finally, Sider can simply reject my request for an explanation as confused: his view is that reference carves at the joints, that this is part of what it is to *be* reference. My request for an explanation of why reference carves at the joints is no more coherent than the request for an explanation of why 1 is the successor of 0: that is just *what it is to be* 1. But it will not do for Sider to simply stipulate that reference carves at the joints, for, as Sider himself points out (see B above), reference is conceptually related to a host of other semantic and quasi-semantic notions: aboutness, belief, intention, etc. Facts about reference explain what our beliefs, intentions, and other mental states are "about" and play an important role in explaining our utterances (and other semantically significant actions). To repeat Sider's own example, part of the reason the teacher says "Maiosarus lived thousands of years ago" is that "maiosaurs" refers to maiosaurs. Part of the reason that believing "the mushrooms in the garden are poisonous," combined with

the desire to avoid poisonous things, disposes speakers to avoid eating the mushrooms in the garden is that “mushroom” refers to mushrooms. So by building joint-carving into the very definition of reference, Sider is thereby assuming that the relation that plays this role (partly explaining a whole host of rational behavior, like utterances) is itself joint-carving. But we can ask Sider: why is the relation that plays a role in action-explanation also a joint-carving relation? If reference (both utterance-explanation *and* joint-carving) has so much built into its very essence, why is it possible in the first place? I do not see how Sider can answer this question, other than by arguing *that* reference is possible, because the alternative—massive reference failure or massive referential indeterminacy—would be (as Sider says, borrowing from Jerry Fodor) “the end of the world.”<sup>69</sup> But that is not an explanation of *why* reference is possible.

*E. Physicalism & metaphysical explanation.* Sider might reply that my request for an explanation of the putative semantic fact that reference is a joint-carving semantic relation goes against the physicalist spirit of his system. Sider accepts two principles about metaphysical explanation:

*Completeness.* Every nonfundamental truth holds in virtue of some fundamental truths.

*Purity.* Fundamental truths involve only fundamental notions.

For our purposes, we can take “fundamental truths” to be truths that have no metaphysical explanation (that do not hold in virtue of anything).<sup>70</sup> Sider has an austere

---

<sup>69</sup> Sider 2011, 28.

<sup>70</sup> Sider 2011, 136. “Metaphysical explanation” is a technical term in Sider’s theory. The metaphysical explanation of the fact that *p* takes the form: “*p*” is True iff \_\_\_\_ , where the

physicalist-cum-nominalist conception of the fundamental truths: they are ultimately truths about the values that various physical fields (e.g the electromagnetic field) take at points of spacetime. Sider can reply to my request for an explanation of why reference carves at the joints by saying that this truth holds in virtue of some highly disjunctive, massively complicated physical truths. Truths about reference are not fundamental (not even close, on Sider's view) so (by Completeness) they must be explained (by Purity) in fundamental terms. But this is merely to claim that there is a metaphysical explanation of why reference carves at the joints, not to provide one. So, while I cannot claim that Sider's view entails *there exists no such explanation* of this fact, I think I have demonstrated that Sider has not explained it, and, in particular, has not explained why the austere physicalist fundamental truths will produce an explanation of why the terms in our language come even close to carving at the joints.

The lesson of all of this is that, contra Sider's 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

---

right-hand side gives the "metaphysical truth-conditions" of  $p$ , which are formulated in purely structural (joint carving) terms.

reboot too extreme to contemplate; and I have no idea how I'd try to convince someone who didn't share it.<sup>71</sup>

In addition to assuming all of this, Sider must assume additionally *without explanation* that the terms of our language (alternately, our concepts) carve the world at its metaphysical joints, or something equivalent (e.g. that “reference” is a joint-carving term, or “joint-carving” is; see above). My point is that there is room to drive a wedge between these: we can admit the world has an objective structure but question whether we are capable of limning it.

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 and metaphysical structure fails to explain why our terms carve at the joints. 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 literature. In the next section I respond to some skeptical worries about the explanatory gap objection itself.

## **§7. Objections and Replies**

---

<sup>71</sup> Sider 2011, 18.

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.

*Just more “just more theory”?* Some readers might object that the problem I have just raised is a version of Putnam’s model-theoretic objection to global descriptivist theories of reference. Likewise, my objection to Sider’s joint carving explanation of metaphysical reference might seem analogous to the “more theory” response to certain solutions to Putnam’s argument. The original “more theory” response was Putnam’s reply to causal theories of reference: the causal theorist includes a causal requirement on reference, which Putnam then interprets as another clause in the theory to be interpreted (“just more theory”), applies his model-theoretic argument, and proves that the reference of the new theory (the old theory plus the causal theory of reference) are just as indeterminate as those of the older theory. I agree with Lewis and Sider that the “more theory” response confuses semantics and meta-semantics: the causal-theory of reference is a meta-semantic theory about what reference in theory T consists in. It is not itself a part of theory T, the semantics of which are to be determined by applying the (massively indeterminate) global descriptivist theory. Likewise, when Lewis imposes a “naturalness” requirement on reference, or Sider a “joint carving” (or his preferred term, “structural”) requirement, it is inappropriate to then simply apply global descriptivism and generate deviant interpretations.<sup>72</sup>

---

<sup>72</sup> Sider puts the point well: “the constraint is not that ‘predicates stand for natural properties and relations’ must come out true on a correct interpretation; it is rather, and

But that is not what I am doing. I am *not* claiming (as Putnam does, *vis à vis* the causal theorist) that Sider's theory, which includes as a premise that *ceterus paribus* we refer to what is structural, fails to constrain the interpretation of the theory, fails to eliminate "unintended" models as referents of our terms. I am pointing out that putative explanations of *why* reference carves at the joints are not explanations. They are not explanations because they assume that reference does carve at the joints, that the reference relation is an assignment-relation (of semantic values to terms in our language) that *ceterus paribus* maps terms in our language to structural items. My claim is not that, once adopted, Sider's theory fails to constrain reference (Putnam's response). My claim is that Sider's theory fails to explain *why* reference is so constrained. Another way to see that my point is fundamentally different is that Putnam's "more theory" response could be raised to *any theory of reference*: this is just more theory. But at no point have I claimed that Sider's theory of reference fails to explain metaphysical term reference because it is "just more metaphysics." What I am looking for from the substantive metaphysicians is a *metaphysical* explanation of metaphysical reference. That is what, I have argued, is lacking in substantive meta-metaphysical theories up until now.

*Not just a problem for metaphysics.* Another likely source of resistance to my argument in this paper is that it proves too much. The same problem arises in mathematics and physics, it might be argued. Why do our mathematical terms refer to mathematical objects? All of the same problems will arise for the explanatory strategies canvassed in

---

more simply, that predicates must stand for natural properties and relations in a correct interpretation" (2011, 32).

Section 5. While the dialectical situation is somewhat different with natural-scientific and natural-kind terms, arguably, many of the same problems will arise. So this is only a problem for metaphysics insofar as it is a problem for mathematics and the natural sciences. Metaphysics is in no special danger.

But we do have explanations of why our mathematical and natural-scientific terms refer: metaphysical explanations.<sup>73</sup> Setting aside mathematics for the moment and focusing on natural-scientific terms, I think we do have an explanation of why these terms refer, which are meta-semantic theories couched in the terms of joint-carving. These terms refer because there are objective joints in nature and our terms *ceteris paribus* carve at the joints. My point in this paper has been that we lack an explanation of why this is the case: we lack an explanation of why reference carves at the joints. That is to say, the meta-semantics of joint carving answers the question of natural kind term reference. Because I think of meta-semantics as part of metaphysics (the part that gives metaphysical explanations of semantic facts) this means that metaphysics (meta-semantics) can explain the reference of natural kind terms, but cannot itself explain a key explanans in that explanation: that reference carves at the joints. There is no science that stands to metaphysics as metaphysics stands to the natural sciences, no science that would back-stop metaphysical term reference, so metaphysics must look after itself. Since, I have argued, the substantive conception of metaphysics has so far failed to do so, I take this to be a problem for substantive metaphysics. In other words, we have (the

---

<sup>73</sup> Some would prefer to call these “meta-semantic” explanations but I think of meta-semantics as that part of metaphysics that looks for metaphysical explanations of semantic facts.

rudiments of) a metaphysical explanation of the semantic possibility of natural science (how reference in natural-science is possible). What is lacking is a metaphysical explanation of the semantic possibility of *metaphysics itself*.

More generally, mathematics and natural sciences 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*. Metaphysics cannot appeal to anything like this history of successes, so it stands in a different dialectical position with respect to its 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 – and, being able neither to explain why our reference carves at the joints nor cite a long and successful history, it has no reply.

## §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 mere 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 other sciences: that it is semantically defective, and this is why it has never entered upon the secure path of a science.<sup>74</sup>

In Section Two I pointed out several reasons 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*

---

<sup>74</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).

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 offer a resounding: No!

### **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/>.
- Blatti, S. & Lapointe, S. (Eds.). (2016). *Ontology After Carnap*. Oxford: Oxford University Press.
- 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.
- Fine, K. 2007. *Semantic Relationism*. London: Blackwell.
- , 2017a. A Theory of Truthmaker Content I: Conjunction, Disjunction, and Negation. *Journal of Philosophical Logic* 46(6): 62–674.
- , 2017b. A Theory of Truthmaker Content II: Subject-Matter, Common Content, Remainder, and Ground. *Journal of Philosophical Logic* 46(6): 675–702.
- 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.
- 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.
- 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.
- 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.
- Russell, G. 2008. *Truth in Virtue of Meaning: A Defense of the Analytic/Synthetic Distinction*. 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.
- 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.
- 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.
- 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.
- 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.

