

Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy

DAVID PLUNKETT

Dartmouth College, USA

(Received 23 January 2015; accepted 21 July 2015)

ABSTRACT *This paper is about philosophical disputes where the literal content of what speakers communicate concerns such object-level issues as ground, supervenience, or real definition. It is tempting to think that such disputes straightforwardly express disagreements about these topics. In contrast to this, I suggest that, in many such cases, the disagreement that is expressed is actually one about which concepts should be employed. I make this case as follows. First, I look at non-philosophical, everyday disputes where a speaker employs (often without awareness of doing so) a metalinguistic usage of a term. This is where a speaker uses a term (rather than mentions it) to express a view about the meaning of that term, or, relatedly, how to correctly use that term. A metalinguistic negotiation is a metalinguistic dispute that concerns a normative issue about what a word should mean, or, similarly, about how it should be used, rather than the descriptive issue about what it does mean. I argue that the same evidence that supports thinking that certain ordinary disputes are metalinguistic negotiations also supports thinking that some (perhaps many) philosophical disputes are. I then explore some of the methodological upshots of this understanding of philosophical disputes.*

Introduction

Much of philosophical inquiry ultimately aims at better understanding some part of reality, rather than our representations of it. Consider questions about what grounds what, or what supervenes on what, or about real definitions. Inquiry into such issues is not about us or our representations of things, or at least not directly so. Rather, it is about the *things themselves*—e.g. what they are grounded in, what they supervene on, or about what they really are. For

Correspondence Address: David Plunkett, Philosophy Department, Dartmouth College, 6035 Thornton Hall, Hanover, New Hampshire 03755, USA. Email: david.plunkett@dartmouth.edu

example, when one wants to know what free will is grounded in, or what it supervenes on, or what it really is, one is not asking questions about the meaning of the word ‘free will’ or about our concept FREE WILL.¹ Rather, one is asking about the thing itself: that is, one is asking about *free will*. The same point can be made if we substituted in other paradigmatic targets for philosophical inquiry—e.g. knowledge, justification, evidence, morality, meaning, justice, truth, beauty, normativity, value, the self, or consciousness.²

This point follows straightforwardly from thinking about what ground, supervenience, and real definition amount to. There are, of course, many important philosophical debates about how to understand each of these three things. But, for my purposes, we can abstract away from the details of such debates and paint in broad brushstrokes. To appreciate the kind of things I have in mind here, we can think of things as follows: (a) ‘ground’ refers to a constitutive (rather than causal) explanatory relation between facts, (b) ‘supervenience’ can be defined in terms of the following: the A facts supervene on the B facts when there cannot be a difference in the A facts without a difference in the B facts, and (c) ‘real definition’ concerns a definition of what something really is (which, on some views, might centrally involve giving an account of its essence).³ The point here is this: these are all *object-level* issues, rather than representational-level ones. So, when philosophers want to know about these kinds of issues, their inquiry isn’t about our representations of reality. Rather, it is about reality itself.

This fact about the aim of much philosophical inquiry—namely, that it often ultimately aims at better understanding some part of reality, rather than our representations of it—is a crucial one to keep in mind when thinking about the nature of philosophy and its methodology. Perhaps in part because of this fact, this picture also informs the default interpretation that many philosophers have of *communication* in philosophy. In short, philosophers frequently interpret what is going on in philosophical disputes in a way that straightforwardly aligns with the

¹Throughout this work, I use small caps to name concepts, italics to introduce terminology or for rhetorical stress, single quotes to mention linguistic expressions, and double quotes for quoting other authors, “scare quoting”, simultaneous use and mention, and other informal uses.

²Of course, some objects of philosophical inquiry—including some of the ones that I just canvassed above—might turn out to be judgment-dependent in a way that slightly complicates this picture. For example, according to some theories of morality, moral facts are grounded in facts about our moral judgments. Suppose that such a theory turns out to be true. In such a case, inquiry into morality would need to involve inquiry into facts about our moral judgments. However, even if this is true, the main point above still holds: when investigating morality and its grounds, our inquiry is ultimately aimed at the thing itself—namely, *morality*—rather than our representations of that thing, e.g. our word ‘morality’ or our concept MORALITY. In such a case, it is *because* of facts about what morality is grounded in that we need to be interested in a certain class of our judgments (namely, moral judgments). But our ultimate object of inquiry remains the same: namely, in this case, morality and its grounds.

³For helpful overviews of grounding, see Rosen, ‘Metaphysical Dependence’; Fine, ‘Guide To Ground’; and Trogon, ‘An Introduction to Grounding’. For helpful overviews of supervenience, see McLaughlin and Bennett, ‘Supervenience’; and Chalmers, *The Conscious Mind*. For a helpful overview of real definition, see Rosen, ‘Real Definition’.

object-level focus outlined above (rather than a representational-level focus). For example, consider here the question in moral philosophy about whether or not moral facts are fully explained by facts about the promotion of value. Consequentialists say yes, while anti-consequentialists say no. When pushed to say what is going on in *actual linguistic exchanges* between consequentialists and anti-consequentialists, many involved in those disputes (as well as many metaethicists who interpret those disputes) will say this: the exchanges straightforwardly express disagreements about morality, and its explanation. In this paper, I am going to present a framework on which this kind of understanding is sometimes (perhaps often) mistaken. According to this framework, in many philosophical disputes that are seemingly about the grounds, supervenience base, or identity of something—e.g. morality, justice, normativity, etc.—the disagreement that is expressed is not actually about these things. Rather, it is about our ways of representing reality, and, more specifically, about which ways of representing reality we *should* adopt. In slightly more general terms, the normative question is often this: *which concepts we should employ* for the purposes at hand.

Philosophers often overtly and explicitly debate which concepts we should employ, or, more generally, normative and evaluative issues about the tools we are using in our thought and talk. Indeed, these kinds of debate occur throughout the field, in subfields ranging from meta-metaphysics to formal epistemology to applied ethics.⁴ For example, take Sally Haslanger's influential work on race and gender. In much of her work, Haslanger argues that we should replace the concepts that we normally use to talk about social identity (including issues about racial and gender identity) with new ones that she has engineered. Her proposed concepts are supposed to be *better* than the default ones she aims to replace; and, in particular, better for accomplishing the specific philosophical and political aims that Haslanger thinks we should be pursuing. In short, they are concepts that are meant to help us better appreciate (and emphasize) those features of social reality that are important to mark for the pursuit of social justice.⁵

Or, to take another example, Theodore Sider argues in *Writing the Book of the World* that, at least for the purposes of doing metaphysics, we should use those concepts that match up with the objective joints in reality. In rough terms, these will be concepts that pick out those properties that David Lewis describes as 'natural' ones.⁶ Such natural properties (as opposed to gerrymandered ones, such as the property of being a quark-or-elephant) are ones that mark out objective similarities between things, and that have real explanatory import. In pursuing his work in metaphysics, Sider therefore introduces

⁴Alexis Burgess and I provide a critical overview of these kinds of debates in Burgess and Plunkett, 'Conceptual Ethics I'; and Burgess and Plunkett, 'Conceptual Ethics II'.

⁵See Haslanger, 'Gender and Race'; and Haslanger, *Resisting Reality*. It should be noted that Haslanger's views change over time in important ways as her work develops, which is reflected in the essays collected in *Resisting Reality*. What I am describing here is one central strand of much of her thinking.

⁶See Lewis, 'New Work for a Theory of Universals'.

concepts (including his core concept *STRUCTURE*) that he thinks do a better job on this front than the ones we tend to employ in our everyday thinking about reality, and also better than those currently employed by other metaphysicians.⁷

The above examples are cases where philosophers *overtly and explicitly* make normative arguments about what concepts we should employ, or, more generally, normative and evaluative arguments about the tools we are using in our thought and talk. The central argument of this paper, however, is not about such cases. Rather, in this paper, I want to focus on disputes between philosophers that aren't explicitly about such normative or evaluative issues about thought and talk. Instead, in this paper, I am interested in philosophical disputes where the *literal content* of what speakers communicate concerns such object-level issues as ground, supervenience, or real definition. It is tempting to think that such disputes straightforwardly express disagreements about these topics. And, indeed, this is exactly what many philosophers think. In contrast to this, I suggest that, in many such cases, the disagreement that is expressed in the dispute is actually one about which concepts should be employed. This can be so, I claim, even when it clashes significantly with the reflective self-understanding of the parties involved in that dispute.

Some philosophers might be tempted to think the following: if it turns out that most philosophical disputes center on disagreements about word usage or concept choice, then this means we should reject an object-level-oriented understanding of the aims of philosophical inquiry, and, instead, take a view of philosophy where it is really aimed more at understanding things about our representational systems (e.g. our words and concepts). This, however, doesn't follow from my proposal, even if it does turn out to cover what is going on in *most* philosophical disputes (which, as I emphasize below, it very well might not). On my proposal, one's normative views about words and concepts are often based on further views about object-level issues that aren't about representational systems at all; such that disagreements about the former kind of issue between two speakers will usually be evidence for disagreement about the latter kind of issue. Put one way: we often argue about a part of reality via having a normative argument about our words and concepts. So what I want to challenge in this paper is *not* the object-level focus of the aims of philosophical inquiry. Rather, what I want to challenge is a view of philosophical *communication* that is often (sometimes unreflectively) coupled with that understanding of the aims of philosophical inquiry. Part of the upshot of the paper is that, in order to defend an object-level-oriented understanding of the aims of philosophical inquiry, one does *not* need to maintain a view of philosophical communication where that communication is directly about object-level issues.

This paper is broken up into three main parts. In Section I, I explain the core idea of what Tim Sundell and I have dubbed a *metalinguistic negotiation*.⁸

⁷See Sider, *Writing the Book of the World*.

⁸See Plunkett and Sundell, 'Disagreement and the Semantics of Normative and Evaluative Terms'.

In basic terms, a metalinguistic negotiation (or, equivalently, a *normative metalinguistic dispute*) is a dispute in which speakers each use (rather than mention) a term to advocate for a normative view about how that term should be used. Throughout this first section, I focus on relatively prosaic cases from outside of philosophy. These disputes are not ones that explicitly invoke philosophical notions such as supervenience, ground, and essence. The reason for this focus on prosaic cases is straightforward. The goal of this section is to get the main linguistic phenomena squarely in view, and to appreciate some of their important features. This is most easily accomplished with more prosaic cases.

Starting out with these more prosaic cases—and, moreover, working through them with care—also has the following additional benefit: it underscores how metalinguistic negotiations are part of our everyday communication outside of philosophy. This raises a very natural thought: if metalinguistic negotiations are part of our everyday communication, it then makes sense to see whether or not they are also part of our communication in philosophical contexts. I think this natural thought points us in a promising direction, and, in Section II, I argue that it can help illuminate what is going on in some (perhaps many) philosophical disputes.

How widespread metalinguistic negotiation is in philosophical communication is a complicated question. It is one that (for reasons I discuss in what follows) depends both on what turns out to be true about communication more generally (e.g. which theory of the metaphysics of meaning turns out to be correct) as well as on complicated empirical questions. Neither of these are topics that I have much space to address in this paper. What I therefore aim to establish in this section is *not* a claim about how exactly widespread metalinguistic negotiation is in philosophy. Rather, what I aim to establish is this: if we have good reason to think that metalinguistic negotiation is something that goes on in ordinary communication, then we also have good reason to think it is going on in philosophical communication as well. Moreover, there are important features of communication in a philosophical context that might in fact support the idea that metalinguistic negotiation is particularly prevalent in philosophical disputes, relative to disputes of other kinds.

This interpretative proposal matters for two reasons. First, it is a significant result for our self-understanding of what we are doing when engaged in philosophical disputes, including, crucially, disputes that many take to be squarely about object-level issues that have nothing to do with word usage or concept choice. In particular, it helps show the appeal and explanatory power of a kind of interpretative proposal that many philosophers are initially quick to dismiss. Second, it matters for thinking about philosophical methodology; since, if a given philosophical dispute really is a metalinguistic negotiation, then figuring out how to best move forward in that dispute will involve figuring out how to best move forward in a metalinguistic negotiation (and, more specifically, a metalinguistic negotiation that is occurring *in a philosophical context*). In Section III, I reflect on this methodological issue and make some broad suggestions for moving forward.

I. Metalinguistic disputes and metalinguistic negotiations

An important part of what goes on in much communication is this: speakers hold fixed their views about what their words mean (and what their interlocutors words mean), and then use those words to communicate about parts of the world that one speaker (or more) doesn't hold fixed views about. The process, however, can also work in the reverse. That is: we can hold fixed our views about the world around us, and then, in turn, *use* (rather than mention) a word to communicate views about the meaning of that word. These views can either be ones about the descriptive issue of what a word *does* mean or about the normative issue of what it *should* mean.⁹

To illustrate what I have in mind here, let's start with the following case, introduced by Chris Barker in his 2002 paper 'The Dynamics of Vagueness'.

Normally, [1] will be used in order to add to the common ground new information concerning Feynman's height:

[1] Feynman is tall.

But [1] has another mode of use. Imagine that we are at a party. Perhaps Feynman stands before us a short distance away, drinking punch and thinking about dancing; in any case, the exact degree to which Feynman is tall is common knowledge. You ask me what counts as tall in my country. 'Well,' I say, 'around here, ...' and I continue by uttering [1]. This is not a descriptive use in the usual sense. I have not provided any new information about the world, or at least no new information about Feynman's height. In fact, assuming that *tall* means roughly 'having a maximal degree of height greater than a certain contextually supplied standard', I haven't even provided you with any new information about the truth conditions of the word *tall*. All I have done is given you guidance concerning what the prevailing relevant standard for tallness happens to be in our community; in particular, that standard must be no greater than Feynman's maximal degree of height.¹⁰

⁹It should here be noted that, in this paper, I use 'word' in a way that the same word (e.g. 'bank') can mean different things, or express different concepts. This is one standard way of talking in certain parts of philosophy, as well as in certain parts of linguistics. However, there are alternative conventions in philosophy and linguistics as well: including, perhaps most importantly, views on which words are individuated in part by what they mean. I think that the former convention is the more helpful one to use for my purposes of this paper. However, if you prefer this latter convention, then that is okay too. If so, you can rephrase everything I want to say and all my main philosophical points will still hold. For example: rather than talking about the same word meaning two different things, you can talk about two different words (ones that are homophonous, but nonetheless distinct in virtue of their different meanings).

¹⁰Barker, 'The Dynamics of Vagueness', 1–2.

For our purposes here, the core lessons from Barker's case can be glossed as follows. Since Barker puts things above in the first-person, let's use the name 'Chris' to refer to the speaker who utters 'Feynman is tall' in the above exchange. When Chris utters 'Feynman is tall', he is not communicating something about Feynman's height. Instead, he is communicating something about what the contextually supplied standard is around here, and, thus, what 'tall' means in this context (in at least one sense of 'means', even if not others).¹¹ He does so by *using* (rather than mentioning) the word 'tall'. Barker calls this basic way of using a term a 'metalinguistic' one.¹² Drawing on Barker, we can define a *metalinguistic usage* of a term as follows: one where a speaker uses a term (rather than mentions it) to express a view about the meaning of that term, or, relatedly, how to correctly use that term.¹³

Now imagine different speakers both employ a metalinguistic usage of a term, but use that same term in *competing* ways. For example, suppose that Ariana overhears Chris say 'Feynman is tall' in the dialogue above. This leads to the following dialogue, where Ariana utters (2c) in response to Chris uttering (2b).

- (2) (a) So what counts as tall in your country?
 (b) Feynman is tall.
 (c) No way. Feynman is not tall.

The issue here isn't about Feynman's height. Everyone in this dialogue already knows what his height is. Rather, Ariana is here *contesting* Chris's view about how to use 'tall' in this context. In so doing, Ariana and Chris are entering into what we can call a *metalinguistic dispute*.¹⁴

¹¹This caveat in the parentheses here ties into an important point. On one sense of the term 'meaning', we use it to refer to the contextually invariant meaning of a word. This is what, borrowing from Kaplan, we can call the 'character' of a word. See Kaplan, 'Demonstratives'. The speaker here doesn't seem to be communicating about *that* at all. Instead, it seems that he is communicating about what the term means, relative to a context. Sticking with Kaplanian terminology, we could put this in terms of the 'content' of that word. I return to this issue in Section II. For now, it will be important to keep in mind that I am here painting in broad brushstrokes when I talk about 'meaning', since some of the distinctions (such as the one above) aren't crucial for making the initial points I need to make.

¹²Barker, 'The Dynamics of Vagueness'.

¹³For more on this way of defining a 'metalinguistic usage' of a term, see Plunkett and Sundell, 'Disagreement and the Semantics of Normative and Evaluative Terms'; and Plunkett and Sundell, 'Dworkin's Interpretivism'.

¹⁴In Barker, 'The Dynamics of Vagueness', Barker does not consider speakers expressing disagreements via competing metalinguistic usages of terms. Sundell introduces these kinds of disputes in Sundell, 'Disagreements About Taste'; and Sundell, 'Disagreement, Error, and an Alternative to Reference Magnetism'. Barker picks up on the same idea in Barker, 'Negotiating Taste'. Sundell and I discuss these cases at further length in Plunkett and Sundell, 'Disagreement and the Semantics of Normative and Evaluative Terms'; and Plunkett and Sundell, 'Dworkin's Interpretivism'.

There are two parts to what makes something a metalinguistic dispute. There is the part marked by the term ‘metalinguistic’: at least one speaker employs a metalinguistic usage of a term. And then there is the part marked by the term ‘dispute’. For our purposes here, we can define *dispute* as follows: a linguistic exchange that appears to express a disagreement.¹⁵ This is a stipulative definition, which is introduced in order to have a helpful label for the phenomena at hand. It is worth emphasizing that, on this definition, a dispute might or might not turn out to *actually* express a disagreement. Whether it does or not requires further investigation. To repeat, all that matters for qualifying it as a dispute is that (a) it is a linguistic exchange (e.g. this could be two people talking to each other, or two people communicating via writing, etc.) and (b) that this linguistic exchange *appears* to express a disagreement (e.g. to an observer, or to a participant in that exchange).

In turn, we can take a *disagreement* (at least for the relevant kinds of purposes at hand) to be something that involves a kind of *rational conflict* in mental states. This will be the basic view of disagreement that I will also adopt here.¹⁶ The paradigm case here is of two people holding rationally conflicting beliefs, such as, most straightforwardly, when one person believes a proposition that the other person believes the negation of. But, for the purposes at hand, we can also be inclusive here. For example, perhaps other conflicts in mental states (including conflicts in desires, plans, or other kinds of mental states) could constitute a disagreement as well. And perhaps there are beliefs that are also rationally incompatible for reasons that are different than what I have just stated. These further issues will depend on how one thinks about ‘rational incompatibility’, and related issues about how one thinks of propositions and the content of thoughts.¹⁷

The above characterization of disagreement is one where ‘disagreement’ refers to (a) a state (rather than an activity), and (b) involves the attitudes of subjects.¹⁸ When disagreement is understood in this way, it has nothing in particular to do with linguistic exchange, let alone a specific kind of linguistic exchange where one kind of linguistic mechanism is privileged. Instead, on

¹⁵My terminology here, and, indeed, throughout this section, is drawn from Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; and Plunkett and Sundell, ‘Dworkin’s Interpretivism’.

¹⁶This thesis need not be read as a full definition of what disagreement is. It is enough for my purposes here that it is a theoretically helpful characterization of one important feature of disagreement.

¹⁷I discuss this way of thinking about disagreement further in my co-authored work with Sundell. See Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; and Plunkett and Sundell, ‘Dworkin’s Interpretivism’. For helpful connected discussion, see Gibbard, *Thinking How to Live*; and Marques and García-Carpintero, ‘Disagreement about Taste’.

¹⁸In making the distinction between ‘state’ vs. ‘activity’ notions of disagreement, I am here drawing on MacFarlane, *Assessment Sensitivity*. MacFarlane’s distinction draws on Cappelen and Hawthorne, *Relativism and Monadic Truth*, 60–61.

this way of thinking about disagreement, it is natural to think (as, indeed, I think is correct) that speakers can express their disagreements via a range of linguistic mechanisms. In turn, one important question for philosophers of language is to figure out what those mechanisms are and how they work.

Here is one crucial distinction among the variety of ways in which we use language to express our disagreements. One way in which speakers can express a disagreement is via the *literal content* of what is said in an exchange. We can use the term *canonical dispute* to refer to any dispute that centers on the content literally expressed by the speakers.¹⁹ Philosophers often tend to privilege canonical disputes when thinking about how disagreements are expressed (including, crucially, when thinking about their own disputes). But not all disputes are canonical. Some disputes center on content that is *not* communicated via the literal content of what is said, but is instead communicated via a range of other ways—including via such mechanisms as implicature, presupposition, or connotation. Following one standard way of carving things up in the philosophy of language, we can (at least for our purposes at hand) think of the above division as mapping onto something worth calling the distinction between “semantics” and “pragmatics”. In short, we can distinguish between information that is part of the linguistically encoded content of the words we use (the domain of *semantics*), as opposed to information that is communicated in some other way (the domain of *pragmatics*). Some disputes—disputes that center on information that is communicated via pragmatics, rather than semantics—are *non-canonical* disputes.²⁰ And, crucially, non-canonical disputes—of which metalinguistic disputes are one important variety—can also express genuine disagreements.²¹

To see this, return to our above metalinguistic dispute between Ariana and Chris. In this case, it seems that Ariana and Chris disagree: one of them believes a proposition about how to use the term ‘tall’ in this context that the other one believes the negation of. Those beliefs are in rational conflict. Those beliefs aren’t communicated via the literal content of what each says. But they are still nonetheless communicated.

It is important to emphasize that this case is thus different than a case where speakers ‘merely talk past’ one another. In basic terms, cases of speakers ‘merely talking past’ one another are ones where there is a dispute, but that dispute turns out *not* to actually express a disagreement, given the facts about

¹⁹For further discussion, see Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; and Plunkett and Sundell, ‘Dworkin’s Interpretivism’.

²⁰As with the terminology of ‘canonical disputes’, this terminology of ‘non-canonical disputes’ is drawn from Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; and Plunkett and Sundell, ‘Dworkin’s Interpretivism’.

²¹This is one of the central theses that Sundell and I argue for at length in Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; and Plunkett and Sundell, ‘Dworkin’s Interpretivism’.

how the speakers differ in what they mean by their words. Consider the following dialogue:

- (3) (a) Agnes was at the bank this morning.
 (b) No, that is wrong. I was at the bank all morning and Agnes never showed up.

(3) is a linguistic exchange. And let's suppose that, at least initially, both speakers take themselves to mean the same thing by 'bank'. In that case, (3) is a dispute: the linguistic exchange appears (at least to the participants, at least initially) to express a disagreement. But now suppose that, in fact, the speaker of (3a) uses 'bank' to refer to the side of a river and the speaker of (3b) uses 'bank' to refer to a financial institution. In the version of the case I have described, the reason that the speakers initially took there to be a disagreement in the first place had to do with thinking they meant the same thing by 'bank'. From their perspective, if they had known all along they just meant different things by 'bank', there wouldn't have even been the appearance of a disagreement between them. So once that presupposition of shared meaning is defeated, they likely will realize they don't disagree about anything here at all. So they don't, for example, think that one of them has the wrong descriptive understanding of what the word 'bank' *does* mean, nor are they advocating for normative views about what it should mean.²² There thus isn't any sort of rational conflict between the thoughts they express in their dispute, and, hence, their dispute doesn't actually express a disagreement.

Some metalinguistic disputes might turn out to be disputes that don't actually express a disagreement, either for the kinds of reasons I glossed above in the case of a 'mere talking past' or for any number of other reasons. But many metalinguistic disputes—including the one between Ariana and Chris in (2)—are different. In (2) Ariana and Chris *do* disagree about something: they disagree about what the word 'tall' means in this context. They each communicate their beliefs about this topic (beliefs that are in rational conflict with each other) via their competing metalinguistic usages of the term 'tall'. Thus, (2) is not only a dispute. It is also a dispute that actually does express a disagreement. And it is thus different than a case of speakers 'merely talking past' one another.

I.i. Metalinguistic negotiations

As I will define things in what follows, a *metalinguistic negotiation* (or, equivalently, a *normative metalinguistic dispute*) is a specific type of metalinguistic dispute. In basic terms, it is one that centers on normative issues about

²²If you are worried about my way of talking here on which the same word 'bank' means two different things, see my earlier discussion in footnote 9.

what a word *should* mean, or, similarly, about how it *should* be used, rather than the descriptive issue about what it *does* mean or about how it *is* used.²³

Many metalinguistic disputes are *not* metalinguistic negotiations. Consider the above case with Ariana and Chris. A natural thought to have is that the dispute between Ariana and Chris is about a descriptive issue. Consider the plausible gloss that Barker gives of the meaning of the term ‘tall’: “having a maximal degree of height greater than a certain contextually supplied standard”.²⁴ We can certainly fill in the details of the case above such that there is good reason to think that Chris and Ariana both mean something along these lines by the term ‘tall’. Let’s suppose, then, that this is so. In that case, we can draw on Kaplanian terminology and gloss things as follows: there is agreement between them about the *character* of ‘tall’ (it’s contextually invariant meaning) but disagreement about the *content* of ‘tall’ (what the term means, relative to a context).²⁵ If that is the case, it is natural to read them as arguing about a descriptive issue; namely, as Barker puts things, as arguing about ‘the prevailing relevant standard for tallness’.²⁶ This is what we can call a *descriptive metalinguistic dispute*: it is about a descriptive issue about word usage or meaning.

In contrast to this, metalinguistic negotiations are not about this kind of descriptive issue. Indeed, they are not about a descriptive issue at all. Instead, they are about what a word *should* mean, or, similarly, about how it *should* be used. (This is why I use the label of *normative metalinguistic disputes* as another label for *metalinguistic negotiations*).

²³Some philosophers might immediately want to object to this distinction based on the idea that meaning itself is a normative issue. Does Sundell’s and my understanding of what makes metalinguistic negotiations distinct rule out this view? No. Even those drawn to the idea that meaning is normative will need to capture the distinction we are drawing here, between facts about a word’s current meaning or usage vs. facts about what it should mean or how it should be used. If one thinks that meaning is normative, there are still a variety of different ways that this might be done. Here is one way to proceed. If we take on board the idea that meaning is normative, then we should take on board the idea *all* metalinguistic disputes involve a normative issue; in basic terms, an issue about how to properly use a word in the context at hand. But we can then make a distinction among metalinguistic disputes based on *which norms* are appealed to by the speakers in the dispute. Some metalinguistic disputes (those that I earlier glossed as about descriptive issues about meaning) are about how to conform to the norms *internal* to the actual current meaning of ‘tall’; the norms that are set forth by (or constituted by) the current meaning of a term. Other metalinguistic disputes invoke *other* norms (e.g. all-things-considered norms about how to live, or what to do that are the kinds of norms invoked in much discussion in ethics, epistemology, and political philosophy). These latter ones are metalinguistic negotiations.

²⁴Barker, ‘The Dynamics of Vagueness’, 2.

²⁵See Kaplan, ‘Demonstratives’. It should also be stressed here that the core of my points here in what follows hang entirely free from the details of Kaplan’s framework. If you prefer other frameworks, such as those argued for in Lewis, ‘Scorekeeping in a Language Game’; or Stalnaker, ‘Assertion’, that is fine for my purposes in this paper. The Kaplanian terminology is used solely for ease of exposition and in order to help readers get a basic handle on the basic kind of distinction that matters here.

²⁶Barker, ‘The Dynamics of Vagueness’, 2.

To better grasp what a metalinguistic negotiation looks like, let's introduce another example. To keep things relatively simple to start with, stick with the case of a term that is context-sensitive and gradable in a similar way to 'tall'. Take the term 'hard'. 'Hard', like 'tall', is a term that denotes a specific property only when a threshold has been set by the conversational context, or by the parties in the conversation.²⁷

Suppose that two friends, Anna and Dan, have decided to open up a mini-golf course together. Anna and Dan have decided that they will rate the difficulty of each hole on the course. They have decided to use a rating system with three different ratings: easy, intermediate, and hard. Anna has been advocating for a higher bar for which holes should get the label of 'hard'. Her main reason for this is as follows: she wants to have a mini-golf course that caters toward expert golfers. Based on this, she has advocated for the following idea: she wants to allow customers to buy a specially priced ticket that allows them to play just the holes labeled as 'hard'. She worries that if too many holes get labeled as 'hard' that experts won't actually find that challenging, then they will be dismissive of the golf course. Dan, on the other hand, wants a mini-golf course that caters toward beginner golfers, such as young children trying out golf for the first time. He wants them to feel good about their progress, and so wants to make sure to have some runs labeled as 'hard' that they will feel good about being able to complete. Anna and Dan have often argued about their conflict explicitly.

With this background information in mind, now suppose that Anna and Dan find themselves having the following dispute, with Dan uttering (4a) and Anna uttering (4b):

- (4) (a) That hole that we just played together was hard.
 (b) No way. That wasn't hard.

Suppose that Anna and Dan just played that hole together, and that they both are basically equally as skilled at golf. We can also here stipulate that they each played the hole equally well. Given all of that, now ask: what is going on in this dispute? A natural suggestion here is this: they are arguing about how they *should* set the threshold for 'hard' in this context. Anna knows that if they apply the label 'hard' to this hole, then this is going to be tied to a more lax labeling scheme—the kind she wants to avoid, given the kind of business she wants to run. And Dan thinks that if the threshold is set too high, then this will be less good for creating the kind of business that he wants to run.

In this case, there aren't antecedently settled facts about what the term 'hard' means that will settle the disagreement between Anna and Dan about proper usage of 'hard' in this context. This is because they are not just trying to conform their usage of 'hard' to facts about current usage—whether it is their

²⁷Both of these terms are what Chris Kennedy calls 'relative gradable adjectives'. See Kennedy, 'Vagueness and Grammar'.

current usage, of the current usage within some community. Rather, they are arguing about different ways of how they should use the term going forward. The dialogue in (4) is thus an example of a metalinguistic negotiation.

I.ii. Metalinguistic negotiations about character

There are many other important dimensions along which metalinguistic disputes vary from each other. In what follows, I want to focus on one other specific dimension, a dimension that helps bring out the kind of case that I think is crucial for thinking about philosophical disputes. The above cases that I gave of metalinguistic disputes were all ones in which the basic context-invariant meaning of the term in question (which I put in terms of the ‘character’ of the term) was held fixed. In turn, the dispute centered on an issue about how the threshold for the context-sensitive, gradable term—what that threshold actually is, given the current meaning, or, instead, how it should be set. But let’s step back from this to think about meaning more generally. There are different aspects to the meaning of a term, and, in principle, metalinguistic disputes (including metalinguistic negotiations) can occur about any of these aspects.²⁸

With this abstract point in mind, the point I now want to emphasize is this: a metalinguistic dispute can target the most general aspects of the meaning of a term (whatever that amounts to on one’s specific theory of language). Using Kaplanian terminology, a key part of this basic point is this: a metalinguistic dispute (including a metalinguistic negotiation) might target the basic *character* of a term. We can illustrate this with the case of a metalinguistic negotiation.

Consider the following case, originally introduced by Peter Ludlow in his paper ‘Cheap Contextualism’.²⁹ In that paper, Ludlow reports listening to a sports radio show that was having a discussion about a list of the fifty greatest athletes of the twentieth-century, recently published by the magazine *Sports Illustrated*. The racehorse Secretariat is on the list. As part of the show, there was a heated debate about this choice. Simplifying a bit, we can imagine the following dialogue occurring as part of the debate between two speakers:

- (5) (a) Secretariat is one of the fifty greatest athletes of the twentieth-century.
 (b) No, Secretariat is not one of the fifty greatest athletes of the twentieth-century.

²⁸The core point I am making here is a schematic one, and thus doesn’t depend on enumerating what exactly those aspects of meaning are (your views on that will be obviously be tied into your general theory of meaning). Philosophers with a wide range of different views about what aspects of meaning there are can (and should) still all agree that metalinguistic disputes (including metalinguistic negotiations) can target different aspects of meaning.

²⁹See Ludlow, ‘Cheap Contextualism’. Sundell and I also discuss this case at length in Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’. My discussion below closely parallels that discussion.

Now imagine—as, indeed, seems very plausible in this kind of context of a sports radio show—that both speakers are fully aware of all of Secretariat’s virtues as a race horse; including how fast he was, how many races he won, etc. And we can also easily imagine each speaker is aware of the fact that the other speaker is aware of these facts too. Even with mutual awareness of such facts we can still imagine the dispute continuing in a heated way. As part of the dispute, suppose the following dialogue takes place between the speakers:

- (6) (a) Secretariat is an athlete.
 (b) No, Secretariat is not an athlete.

Now suppose that, throughout the conversation, the speaker of (6a) systematically applies the term ‘athlete’ in a way that applies to both human and non-human animals. In contrast, the speaker of (6b) systematically applies the term ‘athlete’ in a way that always excludes non-human animals. This happens in non-defective conditions, with both speakers aware of the relevant empirical facts about the non-human animals (their accomplishments in races, etc.), and, moreover, with both speakers aware of the difference in how they apply the term.

How should we explain what is going on in such a case? Ludlow proposes that what is happening in the Secretariat case is this: each speaker is advocating for their preferred way of using the term ‘athlete’, in the context at hand. Sundell and I think this basic idea is correct. Put in our terminology: Ludlow’s Secretariat case is a metalinguistic negotiation. Importantly, even if ‘athlete’ is ultimately some sort of context-sensitive term, it is of a different kind than ‘tall’ and ‘hard’. More to the point, the issue here doesn’t seem to be about how one should set some context-sensitive threshold (as in the case with Anna and Dan negotiating about how to set the threshold for ‘hard’). Nor is it about how we should fill in parts of the meaning of a shared term that are vague or otherwise yet determined. Those are both things that can be the target of metalinguistic negotiation. But the issue here is about a more fundamental aspect of the term’s meaning: it is about the *character* of the term ‘athlete’. So Ludlow’s Secretariat case is an example of a metalinguistic negotiation that targets the character of a term; that is, the most general meaning of the term.

Let’s flesh out this picture in a bit more detail. Given what each speaker in fact does mean by ‘athlete’, each of them plausibly says something true when uttering (5a) or (5b), respectively. The same goes with (6a) and (6b). But communicating those true beliefs is not the only thing these speakers are doing. In addition to this, each speaker is also advocating for a normative view about how the term ‘athlete’ should be used in this context. This advocacy happens via their metalinguistic usage of the term ‘athlete’. If we think of the meanings of terms as the concepts that they express, then we can put this thought as follows: each speaker has a preferred concept for using in the context in hand, and they are advocating for that view. The disagreement that is

being expressed in the exchange, then, concerns a normative issue about word usage and concept choice: one of them holds a view about how the term ‘athlete’ should be used that the other denies.³⁰

It is possible to hold the normative view that all ways of using the term ‘athlete’ should be used in this kind of context are equally as good. But that is a bad view. To see this, suppose that one person advocated for using the term ‘athlete’ (in this context) to refer strictly to chairs and elephants. That is a worse view than other alternatives, given the kinds of things that that speakers are trying to talk about and coordinate on in this context. So, even if both speakers in (5) and (6) are speaking truly, it very well might be that one (or both) of them is getting it wrong in their views about what concepts should be employed in this context.

The basic evidence that favors this interpretation of what is going on is the following. First, there is pressure to think each speaker in the dispute means different things by the term ‘athlete’; that is, that they each express different concepts by ‘athlete’. The reason for this is straightforward. One important piece of evidence for what a speaker means by a word is her usage of that word in non-defective circumstances; or, in a connected vein, her dispositions about how to use a word in non-defective circumstances.³¹ We have such evidence in the Secretariat case. Indeed, we even have something stronger: the speakers use the terms in divergent ways, given full awareness of the relevant factual information, and given mutual awareness of how the other speaker applies the term. So the case for speakers meaning different things by ‘athlete’ in the Secretariat case is strong.

Second, there is pressure to think the speakers in this dispute are engaged in a dispute that really does express a disagreement, and is not just a ‘mere talking past’. Part of what supports this idea is this: the speakers are engaged in a heated debate, take themselves to be at odds, and continue in their dispute even as their divergent patterns of applying the term ‘athlete’ become clear.

Third, it makes little sense to think that they are engaged in a dispute about what the term *does* mean. To see this, suppose that a linguist who is an expert about word usage came on the air of the radio show and said ‘I am certain that “athlete” *just means* something that excludes non-human animals’. This might settle the issue for someone who thinks that that way the term should be used is one that conforms to current usage. (There is no reason, after all, that one can’t advocate for a conservative view in conceptual ethics, rather than a reforming or revolutionary one, when engaged in a metalinguistic negotiation). But it is easy to imagine that such (purported) information about the current meaning of ‘athlete’ would *not* be taken as dispositive by the speaker of (5a),

³⁰For further discussion, see Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’.

³¹On many views of meaning, such things will be what (partially, or perhaps fully) determine what a given speaker means by a given word. But we don’t even need to appeal to that here. Rather, we need only appeal to these things as evidence for what a speaker means by a word.

who is the fan of Secretariat. In certain more self-reflective cases, the speaker of (5a) might very well *grant* that this is what the term ‘athlete’ does currently mean, but not particularly care. She might want the meaning of the term ‘athlete’ to be one that is not human-centric, regardless of whether it currently is or not.

Suppose that this metalinguistic proposal captures what is going on in the Secretariat case. One might then want to know: why should speakers care about this sort of issue? If they agree on all the descriptive facts about Secretariat and his activity, why care about whether the term ‘athlete’ is used to apply to Secretariat or not? The basic answer is this: even aside from what is analytic to a given term (insofar as anything is), there are many things in a given social-historical setting that are associated with a term. A term like ‘athlete’ has rich cultural and emotional associations. Those associations will often be harder to shake than the specific application-conditions associated with the term itself, or the specific meaning it has at a given time. In the case of the dispute on sports radio, it seems reasonable to think that certain things about praise and fame are at stake: after all, being on such a list in *Sports Illustrated* is no small feat. So, in such a case, it can matter a lot not only which of a range of competing concepts one uses, but, moreover, that this concept is paired with the term ‘athlete’ at the heart of the dispute. Similar points can be made about a range of terms with deep resonance in a given social and cultural setting, including, for example, terms like ‘freedom’, ‘rape’, ‘torture’, or ‘gender’.³² This will be something that will be crucial when I turn to philosophical disputes in the next section.

In order to better appreciate why the kinds of dispute that is going on in the Secretariat case can matter, let’s recall the examples that I gave in the introduction of explicit normative arguments about concept choice (arguments given by Haslanger and Sider). These are arguments in what we can dub *conceptual ethics*. We can use this term to cover normative and evaluative issues about thought and talk, including, centrally, normative issues about which concepts one should use in a given context.³³ Haslanger’s and Sider’s arguments concern this central topic in conceptual ethics. And so does the dispute in the Secretariat case. They all concern questions of what concepts we should use, in a given context.

This basic normative topic is a substantive one. It is substantive because not all concepts are equally good for using in a given context, and so we should be using certain concepts rather than others. This basic thought can be

³²For connected discussion, see Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; Chalmers, ‘Verbal Disputes’; and Haslanger, ‘Gender and Race’.

³³This definition draws on my co-authored work with Alexis Burgess. See Burgess and Plunkett, ‘Conceptual Ethics I’; and Burgess and Plunkett, ‘Conceptual Ethics II’.

accepted by people with a wide range of views about what makes a concept better to use in a given context than others.³⁴ Crucially, the fact that normative issues about concept choice matter does *not* depend on the way in which speakers argue about those issues. Rather, it depends on the content of those issues. Thus, this kind of issue in conceptual ethics remains a substantive one regardless of whether it is argued about via the assertion of literal content (i.e. via the semantics) vs. other aspects of the communication (i.e. via the pragmatics). The Secretariat case shows that we can (and do) argue about normative issues about concept choice in non-explicit ways by employing pragmatic mechanisms for communicating information. In particular, it shows that we can (and do) argue about such issues by employing competing metalinguistic usages of a term.

Before moving on, let me note three final points about metalinguistic negotiations. The first point concerns how I have characterized the disagreement involved in a metalinguistic negotiation. Given how I have characterized things above, a natural thought that one might have is this: isn't the *real* disagreement in the Secretariat case then something about the issues of what kinds of creatures get what praise, fame, etc., and not an issue about word usage and concept choice? And doesn't my characterization of metalinguistic negotiations totally miss this point? These are natural worries to have, and they bring out important points that I will draw on in what follows. Indeed, I think there is something they are getting entirely correct, at least when read in a certain way. But they ultimately misfire as criticisms of the view that I have advocated so far.

To see this, we here need to distinguish two things.³⁵ First, there are the reasons why two speakers enter into a dispute, or, similarly, what reasons they have for pursuing it. Second, there is the immediate topic of disagreement. For example: suppose Anil and Seth are at the bike shop trying to figure out which bike Seth should buy. They are arguing about whether or not a bike at the store looks cool. They are arguing about this *because* they each have a view about *which bike Seth should buy* and so their argument *is a way of arguing about* that broader topic. But the immediate topic of disagreement is more

³⁴For example, one might, like Haslanger, invoke such goods as promoting social justice, or, like Sider, invoke the idea of a concept carving at objective joints at reality. Or one might invoke yet other ideas: that we should use those concepts that allow us to most smoothly gather information about the world, or that allow us to most effectively manipulate it. But we need not settle such issues here. All that one needs, in order to get on board with the idea that this part of conceptual ethics matters, is the basic one sketched above: namely, that not all concepts are equally good for using in a given context, and so we should be using certain concepts rather than others. For a critical overview and discussion of the various kinds of values (or norms) that different philosophers take to explain the normative facts in conceptual ethics, see Burgess and Plunkett, 'Conceptual Ethics II'.

³⁵My discussion of this basic distinction, and its relevance to the kind of issue I am talking about here, draws on Sundell's and my earlier discussion in Plunkett and Sundell, 'Disagreement and the Semantics of Normative and Evaluative Terms'; and Plunkett and Sundell, 'Dworkin's Interpretivism'.

limited: it is about whether the bike looks cool or not. With this in mind, return to my proposal about the Secretariat case. What Sundell and I propose is that the *immediate* topic of disagreement in the Secretariat case is the following normative issue: which concept should be used in this context, by being paired with the term ‘athlete’. The metalinguistic analysis that Sundell and I argue for provides an explanation of how that kind of disagreement is expressed, one that draws on the basic idea of metalinguistic usage from Barker. Of course, there are other things that speakers disagree about that motivate them to engage in that dispute, and those are important to keep in mind in thinking about how to settle the dispute. For example, in the Secretariat case, perhaps the speakers disagree about what kinds of creatures are deserving of certain kinds of social recognition and praise. But that doesn’t mean that these background disagreements are somehow the only *real* disagreements, or that they are the ones that are being immediately expressed in the dispute at hand.³⁶

The second point I want to make concerns the *kinds* of normative judgments involved in a metalinguistic negotiation. The basic point I want to make is this: there can be many different kinds of norms that are appealed to in the course of a metalinguistic negotiation, and, correspondingly, many different kinds of normative judgments that these speakers make. To appreciate this point, let’s draw a distinction between two different sense of the word ‘norm’ (and corresponding senses of ‘normative’). First, we can use ‘norm’ in a thin, purely formal sense as follows: norms are standards that one can use to measure things successfully conforming to or not. There are many examples of this kind of norm: there are norms of fashion, norms of chess, norms of etiquette, etc. This kind of norm can be contrasted with the more robust, full-blooded kind of norm that we invoke when we make *all-things-considered normative judgments*: e.g. when we make judgments in ethics about *all-things-considered* how one should live, or when we make all-things-considered judgments in epistemology about *all-things-considered* how one should form one’s beliefs.³⁷ Now notice the following: when a speaker advocates for a normative view about how a word should be used, or what it should mean, she can appeal to a variety of different kinds of norm. For example, she might appeal to the norms of morality, some set of legal norms, or the norms of some system of etiquette. These different norms differ from each other in a variety of ways, including how robust they are. Corresponding to this, we can distinguish the kinds of

³⁶For further discussion, see Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; and Plunkett and Sundell, ‘Dworkin’s Interpretivism’.

³⁷This distinction between ‘formal’ vs. ‘robust’ understandings of normativity draws on McPherson, ‘Against Quietist Normative Realism’. For related discussion, see Parfit’s discussion of reason in the ‘rule-implying’ vs. ‘reason-implying’ sense in Parfit, *On What Matters* (Volume Two), 267–8; and Copp, ‘Moral Naturalism’. To make vivid the intuitive contrast here, consider the following: one issue is what it takes to conform to the norms set forth by a certain form of etiquette. Another issue is whether I have all-things-considered normative reason to conform to those norms.

normative judgments involved in different metalinguistic negotiations, based on which norms those judgments appeal to. Not all metalinguistic negotiations will involve the same kinds of normative judgment. Tied to this, we can note the following: in order for a metalinguistic negotiation to express a genuine disagreement, there will need to be some normative judgments that speakers are expressing that are in rational conflict with each other. But these need not be all-things-considered normative judgments that invoke the most authoritative, full-blooded sense of normativity available.

The third point that I want to make concerns the nature of concepts. Philosophers have a range of different theories of concepts. So one might very well want to know more about what I mean when I say that different speakers (such as these ones in the Secretariat case) employ different concepts, or advocate for the use of different concepts. The views that I develop here in this paper do not depend on a specific, fully developed view of concepts, and, indeed, are compatible with a wide range of leading theories from the philosophy of mind and philosophy of cognitive science of what concepts are.³⁸ That being said, it is worth saying a bit more about the rough view of concepts that I favor, at least for the purposes in this paper. In basic terms, we can think of the relevant job description of concepts here as this: individual concepts are roughly the equivalent in *mental* representation to what individual words are in *linguistic* representation. In turn, for our purposes here, we can then think of different concepts as being marked by the specific *cognitive significance* that they have for an agent or the specific *cognitive role* that they play.³⁹ In many cases (including the concepts that the speakers each express in the ‘athlete’ case) the different concepts that each speaker is using play a kind of *representational* role, such that they are each categorizing things in the world *in a different way*. These different ways of representing the world each partition the world into ways that are in accordance with that representation, and those that are not. If one wanted, one might then model that *way of representing* as something along the following lines: a function from possible worlds to extensions.⁴⁰ This function is a way of determining the extension of the concept in a given possible world. In this paper, as I said above, I will take the *meaning* of a given word (where this is about the *semantics* of that word) to be given by the concept that it expresses.

³⁸For a helpful overview of different views of the nature of concepts, see Margolis and Laurence, *Concepts*.

³⁹This broad way of thinking about concepts resonates with the basic job description for concepts outlined in Wedgwood, *The Nature of Normativity*; and Chalmers, ‘Verbal Disputes’.

⁴⁰This function can in turn be made more complicated by introducing centered worlds, in addition to possible worlds, and by thinking of them within a two-dimensional framework, of the kind advocated for by David Chalmers and Frank Jackson, among others. See Chalmers, ‘Two-Dimensional Semantics’; Chalmers and Jackson, ‘Conceptual Analysis’; and Jackson, *From Metaphysics to Ethics*. I am sympathetic to both of these complications to this picture I have sketched above. But such complications will not matter for our purposes here.

I.iii. Evidence for a dispute being a metalinguistic negotiation

Earlier in this section, I gave a gloss of what kinds of reasons we have for thinking that the Secretariat case is an example of a metalinguistic negotiation. I now want to generalize from this case, putting things in a more abstract and schematic form.

Ludlow's Secretariat case—or at least the version of the case that I have here introduced and discussed—has a number of important features that support thinking it is a metalinguistic negotiation. Here are some of the most crucial ones, put in schematic form.

- (A) There is good evidence that the linguistic exchange is a *dispute*. That is: there is good evidence that it is a linguistic exchange that appears to express a disagreement.
- (B) There is good evidence that the dispute really does express a disagreement.
- (C) There is good evidence that speakers in the dispute mean different things by (at least) one of the terms in that dispute. There are different things that might provide such evidence for a given term. (What one takes to provide such evidence will of course depend on one's more general commitments in the philosophy of language). I have claimed that one such important piece of evidence is this: the speakers in the dispute have dispositions to systematically apply the same term in divergent ways in the same (non-defective) conditions. And I have claimed that many different philosophers, with a range of different commitments should take that as good evidence, even if there are many other sources of evidence as well.
- (D) There is good evidence that the disagreement expressed in the dispute, insofar as there is one, isn't just about descriptive information about what a word does mean, or how it is used. (This is something that helps provide evidence that it is a *normative* metalinguistic dispute, i.e. a metalinguistic negotiation, and not a *descriptive* metalinguistic dispute). For example: one good piece of evidence here would be that speakers persist in their dispute even when they agree on the facts about a term's current meaning or current use.

It is crucial to emphasize here that these conditions are presented in a highly schematic way. What it will take to show that each condition holds will of course depend on many other issues, including, crucially, figuring out what, in

general, gives us good evidence for variation between speakers in what they mean by a given term.⁴¹

There are different ways that one might try to explain what is going on in a given case where conditions (A)–(D) hold. One important kind strategy is to argue that, despite some evidence that speakers mean different things by their terms, this evidence is outweighed by other evidence. If that is right, then we should conclude that speakers in the dispute do mean the same things by the relevant words. In turn, this will allow us as theorists to hold that the dispute is one that centers on literally expressed content and is, thus, a canonical dispute.

This kind of strategy can be bolstered by a number of different kinds of claim. First, one might grant my claim that dispositions to apply a term are an important piece of evidence here, but claim that, despite initial appearances otherwise, speakers really do converge in their dispositions to apply the term. For example, suppose two people currently differ in what they call ‘law’, but one of them has awareness of further sociological information that the other does not. Given that one of them is currently aware of relevant empirical information that the other is not, the two speakers are in *different* circumstances.⁴² It is thus open to argue that they would actually have the same dispositions about how to apply the term in the *same* circumstances. If there is evidence for that further claim, then this can be used to support the following thesis: the speakers really do mean the same things by the relevant term

⁴¹It should be noted that condition (C) is arguably not necessary for advancing a reading of a case as *some kind* of metalinguistic negotiation. There are arguably some metalinguistic negotiations where speakers use terms to advocate for views about how they should be used, but where the speakers mean the same things by their words. This might be the case, for example, in cases where speakers use the grammatical device of metalinguistic *negation* (in the broad sense discussed by Horn, *A Natural History of Negation*) to engage in metalinguistic negotiation. For example, perhaps that is so in the following case: one speaker says ‘Abe needs some help getting to the party tonight since he is disabled’ and the other speaker responds by saying ‘No. Abe is not *disabled*. He is *differently*-abled.’ (Thanks to Lyndal Grant for helpful discussion on this point and for this kind of example). In such a case, both speakers might mean the same things by the terms they use, but advocate (via using their words, rather than mentioning them) a normative view about which words should be used, in the context at hand. The issue here with these kinds of cases is delicate: whether or not we should think that the speakers mean the same things by their terms here depends partly on (among other things) how we should think about the meaning of thick terms, pejoratives, and slurs, as well as how we should think of metalinguistic negation. Since I am here focusing on the kind of metalinguistic negotiation involved in the Secretariat case, I leave this kind of case—and the important issues it raises—to the side for now. Sundell and I discuss metalinguistic negation at length (in another important kind of connection it has to our views on metalinguistic negotiation) in Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’.

⁴²One might also make similar claims here by drawing on resources developed in two-dimensional semantics (as in Chalmers, ‘Two-Dimensional Semantics’; and Jackson, *From Metaphysics to Ethics*), as well as connected resources in David Braddon-Mitchell’s work on ‘conditional analyses’ of concepts (as developed in Braddon-Mitchell, ‘Qualia and Analytical Conditionals’). In short, this work provides additional (more complex) ways of discriminating between different possible circumstances.

(e.g. ‘law’), after all, and, thus, they are thus best understood as engaging in a canonical dispute.

This kind of strategy can also be bolstered by appealing to general theories of meaning that put pressure on the idea that divergent dispositions of the kind I have cited here are an important enough piece of evidence (in a given case) to warrant thinking that speakers mean different things by their words. For example, consider strong forms of content externalism. In rough terms, on externalist theories, the meanings of our words are determined in part by facts external to the agent, such as facts about expert usage or facts about the objective joints in reality. If those external facts play a large enough role in the determination of the meaning of our terms, then it might be that the kind of dispositions that I have pointed to here aren’t as important as I have claimed, and perhaps we then really should take such speakers to be engaged in a canonical dispute after all.⁴³

This kind of strategy—namely, claiming that the dispute in question really is a canonical dispute, despite the presence kind of evidence I have cited against that claim—can also be bolstered by a range of other views about the meaning of the specific terms in question, or the concepts that they express. For example, one important kind of strategy is to propose an expressivist account of those terms or concepts.⁴⁴ Another is to propose what Ronald Dworkin calls an ‘interpretivist’ account of those concepts.⁴⁵ And yet another is to propose a kind of speaker-relativist account of the meaning of the relevant statements or terms.⁴⁶

An alternative to this strategy is to reject the idea that the dispute in question is a canonical dispute, and to instead hold that it is a non-canonical one. Metalinguistic negotiations are one kind of non-canonical dispute. So arguing that the dispute in question is a metalinguistic negotiation is an instance of this kind of approach.

There are a number of reasons why this metalinguistic approach deserves to be taken seriously. One important reason is the conservation of theoretical resources. As I have shown above, there is good reason to think that we sometimes engage in metalinguistic usage of terms. If so, it makes good sense to

⁴³For some helpful examples of the kinds of content externalism that will help bolster the kind of claim under consideration here, see Sider, *Writing the Book of the World*; and Schroeter and Schroeter, ‘Normative Concepts’.

⁴⁴The idea that expressivism might help bolster this kind of strategy has historically been one important source of motivation for expressivism (and for non-cognitivist proposals more generally). For some of the clearest explanations of why this is so, see the opening parts of both Gibbard, *Wise Choices, Apt Feelings*; and Gibbard, *Thinking How to Live*.

⁴⁵For Dworkin’s clearest statement of how his interpretivism is meant to help with the pursuit of this kind of strategy, see Dworkin, *Justice for Hedgehogs*. For discussion of how Dworkin’s views interact with the kind of metalinguistic proposal I am pursuing in this paper, see Plunkett and Sundell, ‘Dworkin’s Interpretivism’.

⁴⁶See, for example, Egan, ‘Epistemic Modals, Relativism, and Assertion’; and MacFarlane, *Assessment Sensitivity*.

see how much work can be done with this tool. This is particularly so given that basically *everyone* needs this tool in their toolkit, even if strong forms of content- or context-externalism turn out to be true.⁴⁷ Connected to this, the metalinguistic approach doesn't necessitate introducing any new kind of concept, or any new theory of meaning beyond what we already have in a range of mainstream descriptivist approaches to meaning. We don't need, for example, Dworkin's special 'interpretative' concepts, or to endorse a form of expressivism or a form of speaker-relativism. This allows us to sidestep a range of important objections to these other views; including, for example, the Frege-Geach problem for expressivism.

There is no way to establish that a given dispute is a metalinguistic negotiation other than looking carefully at the details of the case, while drawing on our best developed theories of thought and talk. As part of this, we will need to consider the relevant merits and burdens of a range of theories of thought and talk, ranging from expressivism to externalism to speaker-relativism. This is obviously no easy task, and one that, crucially, will itself involve reflection on how widespread it makes sense to think metalinguistic negotiation really is. Appreciating the difficulty of this kind of task is itself an important lesson. And it will be important to keep in mind when thinking about my proposal below about philosophical disputes. For it shows that the question of *how many* philosophical disputes are metalinguistic negotiations is intimately bound up with more general questions about thought and talk (including, for example, ones about the metaphysics of meaning) and also complicated empirical questions about the details of what actually happens in given disputes. But another important lesson can be drawn from the above discussion as well. The lesson is this: if a dispute has features (A)–(D) outlined above, then that gives us some evidence that the dispute is a metalinguistic negotiation. The evidence supporting this analysis might of course be outweighed by other evidence. But there is at least a strong case to be made in favor of analyzing it as a

⁴⁷For further discussion of this point, see Plunkett and Sundell, 'Disagreement and the Semantics of Normative and Evaluative Terms'. It should also here be stressed that endorsement of externalism not only fails to provide a way to avoid accepting the idea that there are some metalinguistic disputes. In addition, by itself, it does little to settle the issue of *how many* metalinguistic disputes there are. It might be true that some forms of externalism (if true) would give us good reason to think there are not that many metalinguistic disputes (e.g. those views, such as the view advocated for in Sider, *Writing the Book of the World*, that make heavy use of so-called 'reference magnets', or the metasemantic view advocated in Schroeter and Schroeter, 'Normative Concepts'). However, others would not yield this result. For example, one might be an externalist and still think there is large amounts of semantic underdetermination in our language; something that will lead us naturally to engage in metalinguistic negotiation in many conversations. This sort of idea is at the heart of Ludlow's work in Ludlow, *Living Words*, in which he argues that what I am here calling 'metalinguistic negotiations' are rampant and a crucial part of *most* communication. In this respect, Ludlow's proposal is a more radical version of the kind of thesis that I am arguing for in this paper. At the same time, Ludlow endorses a strong form of semantic externalism.

metalinguistic negotiation, for the basic kinds of reasons that supported the analysis in the Secretariat case. And, for my main argumentative purposes here in this paper, that is all that I need. For my goal in this paper is not to definitively show that a specific philosophical dispute *is* in fact a metalinguistic negotiation (or even that many of them are), but rather to get us to take seriously the proposal that some (perhaps many) of them are, and then to explore some of the implications of doing so.

I.iv. Mixed cases, and the ways in which disputes evolve

In the next section, I turn my attention away from everyday disputes to those that happen within philosophy in particular. However, before I do so, I want to first underscore an important point about my discussion so far. Thus far, I have presented disputes as being either canonical or non-canonical. For example, I have claimed that a dispute is a metalinguistic negotiation and *not* a canonical dispute. But a natural thought is: can't a given dispute be one that involves the expression of disagreement *both* at the level of semantics and at the level of pragmatics? I think the answer is yes. In such a case, insofar as the disagreement that the dispute *centers* on is the disagreement that is expressed via pragmatic mechanisms, the dispute might still count as non-canonical on the definition that we have been working with here. But, crucially, it will still involve a disagreement that is expressed in the semantics, which, moreover, might be quite important indeed. And there will be some cases (perhaps many) where the disagreement that is expressed in the semantics is equally as important as that which is expressed via pragmatics. Based on this, it makes good sense to think of there being a whole range of mixed cases; in short, cases of disputes that involve disagreements at both the level of literal content and at the level of what is expressed pragmatically.

For ease of exposition, I have not focused on such cases in my discussion above. Instead, I have focused on cases where it is plausible to think that the disputes are not mixed ones in this way; or at least I have described them that way, in order to help get the core linguistic phenomena squarely in view. But the truth of the matter in many disputes might very be much more complex (as, indeed, I think it is). This will be important to keep in mind in what follows when thinking about philosophical disputes. For ease of exposition, I will continue to put my main claims as ones about disputes being a metalinguistic one or not. But the truth might be that the dispute is *partially* a metalinguistic one, or that it involves a significant metalinguistic element. And that would be fine, for my purposes at hand. Such a result would still yield a distinct alternative to an important form of self-interpretation by philosophers that I want to challenge; and, moreover, all of the methodological points I will make at the end would still stand. So, in what follows, I will continue to generally abstract away from this issue, and simply talk about disputes being metalinguistic negotiations (rather than just having a significant metalinguistic element), for ease of exposition.

One final point about mixed cases. So far, I have been talking about mixed cases that (roughly) *at a given time* involve both a disagreement expressed in the semantics and one expressed in the pragmatics. But another important kind of case is this: some disputes (including many philosophical ones) change over time, and involve many different kinds of linguistic exchange. So, for example, a dispute that starts out as a metalinguistic negotiation might turn into a canonical one about conceptual ethics, or even a canonical one about something else. Or something that might start out as a canonical dispute could turn into a metalinguistic one. But that does not mean that the *original kind* of dispute that the dispute initially was is identical to the kind it eventually evolved into. Disputes can change types over time.⁴⁸

II. Interpreting philosophical disputes as metalinguistic negotiations

Let's now turn to philosophical disputes, or, put another way, disputes among philosophers in the context of philosophical argument or inquiry. In particular, I want to focus here on disputes where philosophers literally express views about such object-level issues as ground, supervenience, or real definition. As I said in the introduction, many philosophers involved in such disputes take them to be ones that directly express disagreements about *the things themselves*—e.g. free will, knowledge, etc.—rather than ones about our words and concepts. However, we now are in a very good position to challenge this kind of self-interpretation.

In order to set up my argument, it will be helpful to have a somewhat crisper formulation of the kind of dispute I am interested in, and what kind of interpretation of those disputes I am arguing against. As I said, sometimes philosophers make claims where they explicitly and overtly put forward claims in conceptual ethics. Those are cases where part of the literal content of what a philosopher says involves meta-level representations of words or concepts, as well as normative claims about those words or concepts. The easiest way to grasp the kind of dispute I am interested in is this: I am interested in disputes where speakers do not make that kind of explicit claim in conceptual ethics. Instead, I am interested in disputes where the literal content of what each speaker says concerns object-level issues such as ground, supervenience, or real definition. These are disputes that we might gloss as follows: they are ones that ordinary speakers might gloss as 'about' such object-level issues.

The standard way to provide a non-debunking interpretation of this kind of philosophical dispute is as follows: one takes the dispute to be a canonical one

⁴⁸Depending on how one individuates disputes, the cases that I am concerned with here might be best described as either (a) ones where the *same* dispute turns from one kind of dispute to another kind or (b) ones where one dispute stops and another one starts. For my purposes here, this latter way would also be a totally fine way of describing the situation in the kinds of cases I have described. My points above would then need to be paraphrased, but there would not be any loss of important philosophical content here, given what I am trying to argue.

that centers on disagreements about such object-level issues, disagreements that are expressed via the literal content of what speakers say. And the standard way to provide a debunking interpretation of such a dispute is as follows: one takes the dispute to be what I earlier called a ‘mere talking past’. I want to suggest another non-debunking interpretation of an instance of this kind of dispute. This is that the dispute is a metalinguistic negotiation.

The initial argument in favor of reading the dispute as a metalinguistic negotiation is a simple one. I argued above (in Section I) for the following claim: if a dispute has features (A)–(D) that I outlined above, then this provides good evidence (though, as I have emphasized, still *non-conclusive* evidence) that it is a metalinguistic negotiation. With that in mind, let’s consider philosophical disputes about ground, supervenience, and real definition. Now ask the following: do we have good reason to think that some of the disputes where speakers literally express claims about these topics (or about other philosophical object-level issues) are cases where features (A)–(D) hold? I think the answer here is ‘yes’. And, if that is right, then we have good reason to think that such a dispute might very well be a metalinguistic negotiation. To be clear: in what follows, I do not here want to claim that any particular philosophical dispute is *in fact* a metalinguistic negotiation. And, indeed, I do not need to, in order for my argument to work. Rather, what I want to argue is this: if we have good reason to think that some (perhaps many) prosaic disputes (e.g. the Secretariat case, or the mini-golf case with Anna and Dan) are metalinguistic negotiations, then we also have good reason to think that some (perhaps many) philosophical disputes are metalinguistic negotiations.

To begin to see the case for this, let’s first consider a slightly modified version of Ludlow’s Secretariat case. Imagine that the speakers in the dispute have both been reading some philosophy recently. And, so, as part of their dispute, they deploy a number of terms—including, for example, ‘essence’, ‘ground’, and ‘supervenience’—that aren’t normally part of discussions on sports radio. Would that make a difference to our reasons for thinking that the case is a metalinguistic negotiation? I think not. Consider the following exchange that might occur in this modified version of the dialogue:

- (7) (a) Secretariat is not an athlete. That follows from understanding the basic essence of what athletes are, and, connected to this, understanding what grounds an individual creature being an athlete. This is because part of *what it is* to be an athlete is to be a human. That is just what athletes *are*.
- (b) No, Secretariat is an athlete. You have the wrong view of the nature of athletes here. Whether or not an individual is an athlete is explained by certain kinds of skills she demonstrates. Those skills are ones that can be demonstrated by members of a range of species.

Suppose we keep all of the facts constant from our previous version of the Secretariat case. Does the introduction of more philosophical terminology here, or the move to more philosophically advanced terminology, change the evidence we have for thinking it is a metalinguistic negotiation? I think not. And, indeed, I think there are many examples of metalinguistic negotiations outside of philosophy that involve the use of such terminology. For example, consider here the real-world example of arguments about, in rough terms, whether or not Pluto is really a planet or not. Such arguments often involve people making claims about the *nature of* planets, or claims about what *planets really are*. Many such arguments are ripe for analysis as instances of metalinguistic negotiation.

One might, however, be tempted to think that the introduction of this new terminology in the Secretariat case does make a difference in terms of whether or not we read it as a metalinguistic negotiation. Here is one reason to think this. Suppose that speakers use certain terminology (e.g. 'essence') in a way that shows that these speakers have a disposition to use a given term to refer to *the most explanatorily important kind among a range of similar kinds*. (Such that, for example, to say the 'essence of X' signals that a speaker only calls those things 'X' that she thinks is an explanatorily relevant kind, or a member of such a kind, etc.). Now consider that any story about meaning that takes dispositions to matter for determining meaning will need some account of *which* dispositions matter. If so, perhaps this disposition that is shared by both speakers is really what matters here. If that is right, then what initially seemed to be a case where two different speakers had quite different dispositions to apply the same word in divergent ways (in the same circumstances) is actually better understood as one where speakers have the same dispositions, given the kinds of dispositions that matter here.

To bolster such a case, one might also argue that we need to be more careful about how we think about what it is for two speakers to be in the same circumstances. It could be, for example, that, for our theoretical purposes at hand, the two speakers are actually in what should be counted as distinct circumstances. This is because one of them is in the circumstance of having a further belief about what is explanatorily important that the other one does not. There is thus an important line of argument for thinking that the dispute is actually a canonical one, at least partly in virtue of the use of certain philosophical terminology.

This kind of argument is worth taking seriously. But there is also good reason to think that this is not going to support an across the board thesis about what happens with the introduction of the kind of philosophical terminology I am talking about here. It is not that every time that philosophers use terms such as 'ground' or 'essence' this gives us reason to think that what before looked like highly divergent dispositions (of the right kind, whatever that is) actually turn out to be the same. In short, it is not that every time a philosopher uses such terminology she thereby switches to having dispositions that just target explanatorily relevant kinds in a way that is different from when she

was not using such terminology. So it is not that we will, across the board, have evidence to think that such disputes shift from being a metalinguistic negotiation to a canonical dispute in virtue of the introduction of such philosophical terminology. Indeed, absent such evidence in a given case, we will just have more evidence of the dispute continuing to be a metalinguistic one. This is because the speakers are now showing *more ways* in which they diverge in their application of the term ‘athlete’. And, in many cases, nothing else will have changed.

With this in mind, let’s now turn to an example that is more central to philosophical discussion. Consider the question of whether or not free will is compatible with the truth of determinism (roughly, the thesis that the facts about the way the world is, in combination with the facts about the laws of nature, fully determine all of the facts about the future).⁴⁹ Compatibilists claim that they are compatible. In contrast, incompatibilists claim that they are not. Disputes between compatibilists and incompatibilists often take the form of statements that are literally about the grounds of free will, or, relatedly, the essence of free will. So, for example, we might imagine the following dialogue occurring as part of a dispute between philosophers arguing about compatibilism. (We can simplify things here by thinking of this as a spoken exchange, though the import of much of what I have to say below has to do with what philosophers *write*).

- (8) (a) It is part of the essence of free will that only agents that are capable of fully causing their own actions have free will. This means that their actions cannot be fully determined by events over which they have no control, such as the events of the past.
- (b) No, that is wrong. The nature of free will is such that agents can have free will even if they are not capable of that kind of self-determination. What matters is that they we can hold them morally responsible in the right way, which we can do even if they don’t have that kind of capacity for self-determination that you just described. And that is a good thing, given that we in fact don’t have that kind of self-determination.

Now consider the following. Can we flesh out the details of such a dialogue such that it has the features (A)–(D) that I introduced above (the ones that give us evidence for thinking that the dispute is a metalinguistic negotiation)? And, moreover, is such a way of fleshing out the details one that is representative of how disputes of this kind sometimes happen in philosophy? The answer, I think, is a clear ‘yes’.

⁴⁹For a similar gloss on the central claim of incompatibilism, as well as for a helpful critical overview of the philosophical issues involved in assessing this claim, see McKenna and Coates, ‘Compatibilism’.

Let's go step by step, looking at each feature.

With respect to feature (A): Is there evidence that the linguistic exchange is a *dispute*? That is: is there evidence that it is a linguistic exchange that appears to express a disagreement? Yes.

With respect to feature (B): Do we have good reason to think that the dispute really does express a disagreement? Philosophers involved in such arguments about free will certainly think so, or at least much of the time. Maybe that is wrong. But, for now, for the sake of argument, let's suppose that we want to preserve this thought.

With respect to feature (C): Is there evidence that speakers in the dispute mean different things by (at least) one of the terms in that dispute? I earlier said that one such important piece of evidence is this: speakers in the dispute have dispositions to apply the same term in divergent ways, in non-defective conditions. We can certainly imagine an extended version of the dialogue in (8) where we have evidence that the speakers do have such divergent dispositions. If the speaker of (8a) is convinced that the term 'free will' should be used to talk about something of which incompatibilism is true, and the speaker of (8b) is convinced that the term 'free will' should be used to talk about something of which compatibilism is true, then there is good reason to think that is going to be the case. First, they will differ in which claims of the following form they will be disposed to make: 'free will is grounded in ...', 'the essence of free will is ...', or 'free will supervenes on ...' Second, they will differ in their dispositions to make judgments about which particular agents have free will or not, as well as their judgments about whether there is any free will in a given possible world. In at least certain plausible ways of fleshing out this dialogue, these differences in dispositions about how to apply the term 'free will' don't reflect one-off differences, or differences that only show up in what we might consider to be defective observational circumstances. Rather, they are systematic and widespread, and are expressed in circumstances that we have reason to think are non-defective.

With respect to feature (D): Is there good evidence that the disagreement expressed in the dispute, insofar as there is one, isn't just about descriptive information about what a word does mean, or how it is used? I said earlier that one good piece of evidence here would be this: the speakers persist in arguing with each other even when they agree on the facts about a term's current meaning or current use. So, let's assume for now that there are such facts and that they could be provided to the speakers, and ask the following: would they continue to persist in their dispute? Again, we can certainly imagine the details of the case being filled out such that this is so. Consider research into what everyday people mean by 'free will', or, similarly, what a group of philosophers mean by it. For many philosophers involved in disputes like (8) that is certainly *not* going to be taken as dispositive for the issue in question. Indeed, many will think that such information is not only not dispositive, but, indeed, largely irrelevant.

When we combine what I have just argued above with what I argued in the previous section, we have good reason to hold the following: there are some ways of fleshing out the dialogue in (8) such that it is a metalinguistic negotiation. Moreover, such ways of fleshing out the dialogue are representative of what some disputes in philosophical discussions about free will actually look like. Hence, we have good reason to think that some philosophical disputes that are intuitively ones ‘about free will’, and which many philosophers analyze as canonical disputes, turn out to in fact be metalinguistic negotiations.

Let’s now spell out what this would mean if true. Suppose that (8) is a metalinguistic negotiation. In that case, it would mean that the speaker of (8a) and (8b) could both be speaking *truly* about the things that they are literally saying. That could be so if the speaker of (8a) means something by ‘free will’ that is different than what the speaker of (8b) means by ‘free will’. But they are not talking past each other, even if that is so. Instead, they are each advocating, via competing metalinguistic usages of the term ‘free will’, for a view about what the word ‘free will’ *should* mean in this context. They disagree in their normative views here about how this term should be used, and that disagreement is being expressed in the dispute.

As with the Secretariat case, the issue here is about concept-word pairing: it is about which of a range of competing concepts should be paired with the term ‘free will’ for the context at hand. Why care about this kind of issue in the context of doing philosophy? As David Chalmers helpfully sums it up, one of the core reasons here is this:

Ideal agents might be unaffected by which terms are used for which concepts, but for nonideal agents such as ourselves, the accepted meaning for a key term will make a difference to which concepts are highlighted, which questions can easily be raised, and which associations and inferences are naturally made.⁵⁰

We can expand on this basic idea as follows. As with the term ‘athlete’, there is a lot of important resonance that the term ‘free will’ has. This is resonance that obtains not in virtue of the specific meaning of the term (that is: because of the particular concept it is paired with, in a given context), but rather because of the background social and cultural facts about how that term has been used over time, and what kind of functional role it has played in our philosophical discussions. A similar point applies to many terms that are central to philosophical discussion, e.g. ‘knowledge’, ‘self’, ‘freedom’, ‘morality’, etc. Such terms don’t just have strong cultural importance, but are also ones that show up throughout our explanatory projects in philosophy, in ways that

⁵⁰Chalmers, ‘Verbal Disputes’, 542.

are often interconnected in the following way: how a term is used in one sub-area of philosophical discourse is systematically tied to its uses in many other subareas. Because of such background facts, it can (often) matter a lot to the speakers in (8) not only which concept gets deployed in the context at hand, but also which one gets paired with the term ‘free will’ (as opposed to say, the term ‘free will*’ or some other term). This is because, given the importance of the term ‘free will’ in our philosophical discourse, and the relatively fixed functional role we grant those things called ‘free will’ in our practices, it matters that the concept gets paired with this specific word ‘free will’ as opposed to something else. Again, a similar point holds for lots of other philosophical terms that are central to philosophical discussion, e.g. ‘knowledge’, ‘self’, ‘freedom’, ‘morality’, etc.

Let’s take stock of where we are. I have just argued that (at least one important version of) the dialogue in (8) is a metalinguistic negotiation about ‘free will’, and that this is representative of part of what is going on in actual philosophical disputes that, at the level of literal content, are ones where speakers put forward views about object-level issues. How widespread metalinguistic negotiation is in philosophy is, for the kinds of reasons I explained earlier, a complicated question and not something that I can take on in this paper. But, in approaching this topic, one thing that is important to notice is that nothing I have argued above hangs on specific features about philosophical discussions of *free will* in particular. *Lots* of philosophical disputes have the features that make them plausible candidates for being analyzed as metalinguistic negotiations. This is because lots of philosophical disputes seem to have the features (A)–(D) that I identified in the last section. Consider, for example, the following disputes (characterized here, for the sake of exposition, in terms of what they would be about, were they actually a canonical dispute): disputes between consequentialists and non-consequentialists about the grounds of morality; disputes in metaphysics about the individuation of objects; disputes in the philosophy of law about positivism vs. antipositivism; disputes in political philosophy about the nature of democracy; disputes in the philosophy of language about the grounds of meaning; or disputes about the nature of the self. In disputes where the literal content of what speakers say is about such issues, philosophers often interpret those disputes as canonical ones. But now notice that many disputes where the literal content of what speakers say is about such issues are also ones that have features (A)–(D). Thus, for reasons that parallel what I argued in the ‘free will’ case above, this provides good *prima facie* reason to think that many such philosophical disputes are in fact metalinguistic negotiations.

Importantly, the reason to take this proposal seriously is not just that many philosophical disputes have features (A)–(D). It is also that this proposal has explanatory power, and, moreover, can achieve this without the need to introduce major new machinery into our philosophical toolkit. Here is one

important dimension of the explanatory power of this proposal. Consider that one important aspect of many philosophical disputes—something that is arguably more prevalent in philosophical disputes than in disputes in other branches of inquiry, such as mathematics or biology—is that they seem to have the following combination of two features. First, the philosophers in the dispute seem to agree on a wide range of facts about the world (indeed, in some cases, perhaps what we might be inclined to think are *all* the relevant facts about the world, described without using the term at the heart of the dispute). However, they nonetheless continue to argue with each other at length. In many such disputes, the philosophers involved in the dispute take themselves to be expressing a disagreement in their dispute; and, moreover, this is a view that we as interpreters often want to vindicate. It would be good to have an explanation of what is going on here. My metalinguistic proposal offers a way of explaining what is going on in such philosophical dispute that gives a smooth explanation of what is going on—and one that, moreover, allows us to vindicate our judgment of the presence of disagreement in many of the cases we think it should be vindicated. Moreover, it does so without requiring us to posit anything non-empirical or non-naturalistic in our ontology. Neither does it involve introducing major new machinery into our account of thought and talk. We can instead make due with a tool that we already have in our toolkit anyway, that we need to explain a whole host of everyday, mundane cases of the sort that I introduced in Section I. These facts help support the idea that some (perhaps many) philosophical disputes are in fact metalinguistic negotiations.

It is important to empathize that it might ultimately be that metalinguistic negotiation is not that widespread in philosophy, for the kinds of reasons I sketched earlier (e.g. because a certain foundational semantic theory turns out to be correct). But it is a proposal to be taken seriously, and one that, if wrong, is wrong for more complicated reasons than many philosophers would think from the outset. For example, it is not going to be wrong because it can't explain how the disputes involved express genuine disagreements (it can); or because it conflicts with an object-level-oriented understanding of the aims of philosophical inquiry (it does not); or because those disagreements aren't settled by appropriately objective facts (it can, for reasons I discuss more below).

In the Section III, I discuss some of the methodological implications of this view, assuming it is correct. However, before I do so, I want to do three final things in this section. First, I want to underscore four important points about my interpretative proposal about philosophical disputes. Second, I want to briefly discuss the resonance of my proposal above with Frank Jackson's discussion of free will in *From Metaphysics to Ethics*. This will allow me not only to note an important point of contact with existing discussion, but also underscore part of what is important about my proposal in this paper.

II.i. Four important points about my interpretive proposal

Here are four important points about my interpretative proposal about philosophical disputes.

First, it is important to underscore that I am *not* claiming that there aren't real and important issues about such object-level issues as ground, supervenience, real definition, essence, etc. of the kind that are at the heart of much discussion in contemporary philosophy (and especially contemporary metaphysics). Everything I have said is consistent with thinking that there are important issues here to investigate. To see this, consider the following. Suppose one argued (as I think is correct) that an important part of communication among biologists involves metalinguistic negotiation. (The different meanings of 'species' is a good place to start with such a proposal, as is the different meanings of 'intelligence'). Would that mean that there aren't facts about animals and their behavior to investigate, and then all biological argument is just about normative issues about word and concept choices? Clearly not.

On this point, it is worth again remembering here that, as I emphasized earlier in this paper, the background reasons why speakers engage in a metalinguistic negotiation are standardly bound up with their views about object-level issues (and not representational-level ones about our words and concepts). This stems from the following fact: when a person has a view in conceptual ethics, it is standardly the case that this view depends on a whole host of further normative and descriptive claims that aren't directly about words or concepts at all. This means that when we argue about conceptual ethics (either in a metalinguistic negotiation, or in any other kind of dispute) it is not that this is *all* that the argument is about, in the sense that the argument turns only on a free-standing issue in conceptual ethics. Rather, we often argue about a part of reality via having a normative argument about our words and concepts. Thus, what I am challenging in this paper is *not* the object-level focus of the aims of philosophical inquiry, or, relatedly, the idea that there are significant object-level issues for philosophers to investigate. Rather, what I am challenging is a view of philosophical *communication* that is often (sometimes unreflectively) coupled with an object-level understanding of the aims of philosophical inquiry. In short, what my argument shows is this: in order to defend an object-level-oriented understanding of the aims of philosophical inquiry, one does *not* need to maintain a view of philosophical communication where that communication is directly about object-level issues.

Second, it is important that my proposal is compatible with a wide range of views about the foundations of conceptual ethics. Most importantly, nothing in my proposal suggests that normative issues about what concepts we should use can be settled by voluntary choices that we ourselves make. To the contrary, everything that I have said is compatible with a range of anti-voluntarist views about conceptual ethics: including thinking that issues in conceptual ethics are settled by fully objective, mind-independent normative facts (for example, of

the sort envisioned by non-naturalists in metanormative theory)⁵¹, or that they are largely settled by facts about what the objective joints of reality really are (as on Sider's view about conceptual ethics in metaphysics that I sketched at the start).⁵² This is important to stress for the following reason: many views in meta-metaphysics (or meta-philosophy more generally) that interpret philosophical disputes as ones about concept choice are often coupled with specific views about conceptual ethics. In particular, they are often coupled with some version of the following thesis: the normative facts in conceptual ethics are less objective, or less substantive, than the kinds of facts that we standardly investigate in doing metaphysics, or, more radically, there aren't any normative facts there at all (or normative properties, etc.). An upshot of this would be that if it turns out that something we thought was an object-level dispute is really one about concept choice, then this means that it is about a less objective subject matter than we had previously thought. While this might ultimately turn out to be true, nothing in my proposal entails that result.

Third, I want to briefly underscore an important point about the degree to which my proposal ends up clashing with our ordinary, pre-theoretic way of describing the kinds of philosophical disputes that I have focused on. The class of philosophical disputes that I have focused on are ones where philosophers literally express views about such object-level issues as ground, supervenience, or real definition, and where they do not explicitly make claims in conceptual ethics. These are disputes that we might gloss as follows: they are ones that ordinary speakers might gloss as 'about' such object-level issues. That pre-theoretic understanding of things is importantly different from a developed philosophical interpretation of what is going on in those disputes. The developed philosophical interpretation that is often favored (and which I have argued is sometimes wrong) is that these disputes are canonical ones. Now, one might be inclined to think that our pre-theoretic understanding of the disputes is that they are canonical disputes. But that is not obviously so. Instead, our pre-theoretic understanding of such disputes might in fact be neutral with respect to a reading that treats such disputes as canonical vs. my metalinguistic proposal. After all, my proposal hinges on certain subtle distinctions between the different ways in which we use terms, distinctions that ordinary speakers very well might not be aware of. So it might be that nothing in their pre-theoretic understanding of such disputes is on track, even if my metalinguistic proposal is correct.

This point is particularly important to keep in mind when thinking about linguistic data about belief reports that might seem to count against a metalinguistic analysis of a given dispute. For example, in the Secretariat case, one speaker might say: 'The other speaker believes that Secretariat is an athlete.

⁵¹For a helpful overview of this kind of metanormative view (though not in the context of discussions of conceptual ethics), see Enoch, *Taking Morality Seriously*.

⁵²See Sider, *Writing the Book of the World*.

I disagree with that.’ That might seem to straightforwardly favor reading that dispute as a canonical one, and count significantly against a metalinguistic proposal of the kind that I have put on the table. However, I think that things are more complicated here. This is for the kind of reason that I have sketched above: namely, we need to distinguish between the pre-theoretic understanding of something being ‘about’ object-level issues (which is likely what might be reported in the belief report) and developed philosophical or linguistic theories of what is going on in that dispute (something that it would be surprising if many ordinary speakers in fact had views about). Moreover, insofar as speakers do have views on this latter topic, it very well might be that they are just mistaken in their own self-interpretation. Indeed, if my argument in this paper is right, then that will be true for many philosophers.⁵³

I have just emphasized that my proposal might be consistent with our pre-theoretic understanding of what is going on in a given dispute when we say it is ‘about’ object-level issues. At the same time, it should also be stressed that if my proposal does involve rejecting our pre-theoretic understanding of these disputes, then that is totally fine with me. Our pre-theoretic understanding might simply be mistaken. After all, why should we expect that our pre-theoretic understanding of how communication works gets it entirely correct? We don’t expect that our pre-theoretic understanding of physics gets it right. So why should we expect it is different here, with respect to how communication works in philosophy, or, indeed, how it works in general?⁵⁴

The fourth point I want to make concerns an aspect of my proposal that isn’t about disputes at all. I have put my interpretative proposal in terms of a theory of what is going on in philosophical *disputes*. But notice that part of what makes this proposal viable has nothing at all to do with what goes on in disputes, nor even with what is going on in linguistic communication. Rather, it has to do with a view about what is going on when someone is engaged in inquiry.

Consider here Peter Railton’s work on ethics and morality. Railton explicitly gives *reforming* definitions of many of the core terms he employs in discussing these topics, including, crucially, the terms ‘non-moral goodness’ and ‘moral rightness’. His reason for doing so is based on the following general idea:

⁵³For further discussion of the basic idea in this above paragraph, see Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; Plunkett and Sundell, ‘Dworkin’s Interpretivism’; and Plunkett and Sundell, ‘Antipositivist Arguments from Legal Thought and Talk’.

⁵⁴For further discussion of these issues about the way in which this kind of metalinguistic proposal relates to our pre-theoretic views about what is going on in disputes, as well as critical discussion of the extent to which that matters, see Plunkett and Sundell, ‘Disagreement and the Semantics of Normative and Evaluative Terms’; Plunkett and Sundell, ‘Dworkin’s Interpretivism’; and especially the third section of Plunkett and Sundell, ‘Antipositivist Arguments from Legal Thought and Talk’.

[T]he drawing up of definitions is part of theory-construction, and so is to be assessed by asking (1) whether the analyses given satisfy appropriate constraints of intelligibility and function, and (2) whether the terms as analyzed contribute to the formulation and testing of worthwhile theories.⁵⁵

Railton claims that his reforming definitions meet these tests. In somewhat more specific terms, he argues that these definitions help contribute to a viable theory of the thing he here wants to explain: roughly, what morality is, and how it is related to the kinds of naturalistic facts studied by scientists. Railton's reforming definitions aren't meant to be analyses of our current concepts. Rather, they are meant to be ways of *improving* on our current concepts for the purposes of doing serious explanatory work. In introducing his reforming definitions, Railton sees himself as doing something parallel in philosophy to what we frequently do in the sciences: namely, refine how we use our terms to better explain things.

A core aspect of Railton's thought here can be put as follows: an important part of inquiry involves trying to figure out whether or not one is using the best representational tools possible for the job at hand. I think that is exactly right. Part of what supports my metalinguistic interpretation of philosophical disputes is the idea that we as philosophers are often engaged in an ongoing inquiry in conceptual ethics about which concepts we should use (which often includes thinking about how we should *modify* the ones we currently use). Some philosophers (such as Railton, or Sider or Haslanger) might be quite explicit about that. But even those who are not are often still engaged in making tacit choices about issues in conceptual ethics. Part of what makes the metalinguistic interpretation I have put forward attractive is this background fact about the nature of inquiry in philosophy: namely, that it often involves engagement with the precise topic of metalinguistic negotiation (namely, conceptual ethics), even when that engagement isn't self-conscious, and even when it isn't communicated to others.

II.ii. An illustrative example

I now want to briefly discuss the resonance of my interpretive proposal about what is going on in some (perhaps many) philosophical disputes with Frank Jackson's discussion of free will in *From Metaphysics to Ethics*. As I said above, this will allow me not only to note an important point of contact with existing discussion, but also underscore part of what is important about my proposal in this paper.

In from *Metaphysics to Ethics*, Jackson argues that the folk concept FREE WILL—which, roughly, we can here take to be the main one that most people

⁵⁵Railton, 'Moral Realism', 204.

regularly employ in non-philosophical settings—picks out something that can't obtain if determinism is true; such that, in short, if determinism is true in a given world there is nothing in the extension of that concept in that world. However, Jackson argues that compatibilists give us good reason to replace this ordinary concept (that yields this incompatibilist result) with one that yields a compatibilist result. This, Jackson argues, will involve a change of subject from the original pre-theoretical one that we were talking about when we originally used the term 'free will'; that is, the new thing picked out by our replacement concept (which we might very well henceforth express by the same term 'free will') is something different than the thing we were initially picking out by our folk concept. Yet, Jackson claims that this change is well-motivated. This is because our new concept picks out something that is *more philosophically important* than the things picked out by the folk concept. That is: even if that thing is no longer free will, strictly speaking (which it would not be if we hold that free will is the thing picked out by our folk concept), it is still the *more philosophically important thing* to be talking about. And that, in short, is what should ultimately determine our choice of concepts when doing philosophy. As he puts it:

[C]ompatibilists do, it seems to me, show, first, that the folk concept of free action involves a potentially unstable attempt to find a middle way between the random and the determined, second, that the folk conception is nowhere instantiated, and, third, that a compatibilist substitute does all we legitimately require of the concept of free action. It is hard to see how we could better motivate a limited change of subject.⁵⁶

I want to note two things about Jackson's discussion here. The first is that Jackson himself is making an explicit argument in conceptual ethics, an argument of the same basic kind that Haslanger, Sider, and Railton give in the examples at the start of this paper. The second thing is that Jackson argues that part of the reason for his view in conceptual ethics is what compatibilists before him have argued.

Now consider that many compatibilists do not themselves make explicit claims in conceptual ethics at the level of the literal content of what they argue. Moreover, many of them would *not* interpret what they are doing as offering an argument in conceptual ethics, aimed at motivating a change in subject by replacing our folk concept with a new one. And, indeed, Jackson does not say that they all (or mostly) see themselves in that way. Rather, he just claims that this is what they have successfully argued. One way of thinking about my metalinguistic proposal of free will debates is this: it offers part of an explanation of how compatibilist philosophers managed to argue (successfully, in Jackson's view) for a view in conceptual ethics, even though many

⁵⁶Jackson, *From Metaphysics to Ethics*, 45.

of them didn't make these arguments explicitly, and even though many of them thought they were doing something else. Part of how they did this was by consistently employing a metalinguistic usage of the term 'free will' over time, and thereby showing us the (purported) benefits of pairing the word 'free will' with a new concept.

This underscores part of what is important about my argument in this paper. If my interpretative proposal is right, it illuminates something that many philosophers think (or at least suspect) is going on in philosophical discussion, and, moreover, offers an explanation of *how it happens*. Moreover, the explanation is one that sits nicely with a range of further views one might have in the philosophy of mind and language. Many philosophers have held that a significant part of communication in philosophy involves arguments about concept choice and deployment; and many have thought that this kind of argument about concepts goes on in disputes that, at the level of literal content, aren't about concepts at all. For example, these broad ideas show up, in one form or another, in Charles Stevenson's classic paper 'Persuasive Definitions', Peter Ludlow's recent work, and parts of Robert Brandom's work.⁵⁷ Part of what makes my interpretative proposal different from many such proposals is that my proposal rests on fairly ecumenical starting points about the nature of thought and talk. These are starting points that a range of philosophers can (and, I think, should) get on board with (even if it takes them in a somewhat unfamiliar direction at the end of the day in terms of their overall views about thought and talk). This is part of what makes my argument in this paper a significant one. In short, it puts the metalinguistic option for analyzing philosophical disputes squarely on the map, and it does so by relying only on relatively uncontroversial assumptions.

III. Methodological upshots

In this paper, I have put forward a framework on which some (perhaps many) philosophical disputes—and, in particular, ones that appear to many to be canonical disputes about object-level issues such as ground, supervenience, or real definition—are best understood as metalinguistic negotiations. I have argued that such a framework is worth taking seriously, and, indeed, might illuminate what is going on in important philosophical disputes. One reason that this argument matters is straightforward: it matters for the contribution it makes to develop a reflective self-understanding of what we are doing when we are doing philosophy.

A second reason it matters is because it raises interesting methodological issues. Suppose a given philosophical dispute really is a metalinguistic negotiation. If so then figuring out how to best move forward in that dispute will involve figuring out to how to best move forward in a metalinguistic

⁵⁷See Stevenson, 'Persuasive Definitions'; Ludlow, *Living Words*; and Brandom, *Making it Explicit*.

negotiation (and, more specifically, a metalinguistic negotiation that is occurring *in a philosophical context*). So, then, how should we move forward in such a dispute? In this final part of this paper, I reflect on this topic. Following this, in the conclusion (Section IV), I then briefly discuss an important general lesson of my argument in this paper.

III.i. Two dimensions to successful philosophical thought

Suppose that a given seemingly object-level dispute among philosophers is really a metalinguistic negotiation. It is important to keep in mind that there are then two distinct issues at play here, issues that are often helpful to pull apart and tackle independently. The first issue is whether the literal content of what these philosophers say is correct (e.g. by being true). The second issue is whether or not the concepts we are using in making those claims are ones we should be using, for this context. These are two distinct issues, and we would often do well to tackle them independently. In one respect, we could reach this conclusion by just reflecting on the different dimensions of what makes thought successful in the first place: namely, that one dimension of success involves truth or correctness, and another involves using the right concepts. This shows that we should always be sensitive to both of these issues, whenever we are engaged in serious philosophical inquiry (or, indeed, serious inquiry of any kind). But notice that the issue here is more pressing when the dispute one is actively engaged in is a metalinguistic negotiation. In a metalinguistic negotiation, the question of what concepts one should use is front and center: it is the immediate locus of the disagreement expressed in the dispute. And so it will often be helpful to keep that issue squarely in view when engaged in a metalinguistic negotiation, and avoid thinking that establishing the truth of one's claims is enough.⁵⁸

III.ii. Conceptual analysis, descriptive linguistics, and experimental philosophy

When philosophers are engaged in a metalinguistic negotiation, they are arguing about how they should use their words in a given context. In many cases, such as in Ludlow's Secretariat case, that will involve arguing about which of a range of competing concepts should be employed. As Ludlow's Secretariat

⁵⁸For a forceful argument on behalf of this kind of thought, see Sider, *Writing the Book of the World*. For a helpful case study of how it can be important to keep these two different dimensions of successful thought both squarely in view, in a context where philosophers tend to focus just on questions about truth, see Sundell's discussion in Sundell, 'Eligibility and Ideology in the Vat' of how metaphysical realists should best respond to Putnam's famous brain-in-a-vat argument from Putnam, *Reason, Truth, and History*.

case makes vivid, when this kind of normative issue is on the table, it is unlikely that discovering descriptive facts about what the speakers do mean by their words—or, put another way, how the concepts that they express by those words are best analyzed—will settle what's at stake in the dispute. Both speakers might know full well what the other speaker means by her words. The issue is that at least one of them thinks that the other is *making a mistake* to use the relevant words that way. More generally, once a speaker appreciates that the issue on the table is a normative one about how we should use our words, a speaker should also be hesitant to conclude that a good way to settle the dispute is to analyze the meaning of her *own* words used in that dispute. The facts about how she is using her words can't themselves establish how she *should* be using them. Similarly, unless she has good antecedent reason to defer to the usage of others outside of the dispute, settling on the facts about how others outside of the dispute use their words will also be of limited relevance to settling the normative question on the table.

This is not to say that this kind of descriptive information will be useless here, or totally beside the point. Indeed, it is hard to see how it could be *totally* besides the point. For example, suppose one thinks that a current concept is defective and should be replaced by a new one. Presumably we need some kind of grip on what a current concept is in order to judge that it is defective. Similarly, if we don't have a good understanding of what our new concept amounts to, why should we be so confident that the new one will be better? These kinds of thoughts point to an important role that conceptual analysis can play in a metalinguistic negotiation, and, indeed, in thinking about conceptual ethics more generally. But the point I want to make here isn't that such descriptive facts aren't relevant. Rather, it is just that these particular descriptive facts about our words or concepts won't fully settle the normative issues on the table. Put another way, this means the following: figuring out from conceptual analysis, descriptive linguistics, or experimental philosophy what people mean by their words (including understanding the content of the concepts they express by those words) won't by itself settle normative issues about conceptual ethics. Thus, they won't fully settle the kind of issue that is being argued about in a metalinguistic negotiation.

Importantly, this result dovetails with a thought that follows naturally from the kind of prominent interpretation of the kind of philosophical disputes that I have been challenging here: namely, the interpretation that takes them to be canonical disputes centered on object-level issues. When that interpretation is right, it also has the result that conceptual analysis, descriptive linguistics, or experimental philosophy won't fully settle things. That is because, as I emphasized at the start of this paper, when we are talking about object-level issues (such as the nature of free will) we are not asking questions about our words or our concepts. We are rather asking questions about the things those concepts pick out—e.g. *free will* itself.

But notice a difference here. Some philosophers have thought (for reasons that I am sympathetic to) that, in canonical disputes, conceptual analysis of the *shared* concepts that both speakers express in the dispute can play an important role in helping speakers better understand the topic of discussion. In basic terms: if we are having an argument about something that falls under a given concept, better understanding that concept can help illuminate what must be the case for something to count as falling under that concept in the first place. It can thus help us in figuring out when someone is *changing the topic* of the dispute, and thus help us keep clear about our subject matter.⁵⁹ If that kind of thought is correct, it shows an important role for conceptual analysis to play. But it is of *less* relevance in a metalinguistic negotiation, even by the lights of this kind of proposal. In a metalinguistic negotiation, speakers might have a background set of shared concepts that makes it possible for their thoughts to be rationally incompatible. But—in at least some metalinguistic negotiations, such as the Secretariat case—the speakers are expressing *different* concepts with the words used in the dispute. Thus analyzing *those* concepts is going to be of more limited relevance in getting clear on what the topic of the dispute really is, or in fixing the subject matter. In metalinguistic negotiations like the Secretariat case, speakers are employing different concepts; ones that will often be picking out *different* things in the world, things with different grounds, essences, real definitions, and (often) different supervenience bases. So one issue that is on the table here—an issue they argue about via arguing about which concepts they should use—is which thing we should be talking about here, for the purposes at hand. What fixes the subject matter of their dispute is thus not something that can be discovered via the analysis of the words they use in that dispute.

III.iii. Paraphrase, and moving to background issues

Once one realizes that one is engaged in a metalinguistic negotiation, it can often be helpful to move to having a *canonical dispute* about the issues in conceptual ethics that the negotiation is about. In some contexts, such a move might come at a big cost. For example, suppose one is engaged in a political campaign in America that involves a metalinguistic negotiation about the word ‘freedom’. In such a case, it might be a strategic disaster for a political candidate to say something such as (9b) in response to (9a):

⁵⁹These kinds of thoughts form an important part of Frank Jackson’s extended discussion of the relevance of conceptual analysis to philosophy in Jackson, *From Metaphysics to Ethics*. At the same time, it should be emphasized that Jackson also defends other theses about the relevance of conceptual analysis to philosophy that I am not discussing here, but to which I am very sympathetic. For connected discussion about the import of conceptual analysis to philosophical inquiry, see Chalmers and Jackson, ‘Conceptual Analysis and Reductive Explanation’; and Thomasson, ‘Answerable and Unanswerable Questions’.

- (9) (a) My proposed policy does a better job of promoting freedom than the policy proposed by my opponent.
- (b) Look, I grant that what you are saying is true, given our current use of the word ‘freedom’. So you are, strictly speaking, making true claims about freedom. But what I am proposing is that we switch to using a new concept, call it FREEDOM*, and then expressing it with our existing term ‘freedom’ moving forward. But, if you want, we can just have an argument now about which is the concept we should use: FREEDOM (our current concept), which we can for now call ‘freedom’, or my proposed concept FREEDOM*, which, for now, we can express with the term ‘freedom*’.

The basic reason it can be a strategic disaster to utter (9b) is this: given the rich positive cultural associations with the term ‘freedom’ in America, a politician involved in a political campaign will want those associations tied to her own claims. If she proposes to give up the term ‘freedom’ temporarily—or even if she just makes clear that she is not using the existing notion of freedom that people are familiar with—she risks distancing her views from some of those associations. In short: paraphrase can be politically costly for her, given her aims.⁶⁰

The above kind of paraphrase can be costly in philosophy too, especially with central terms such as ‘knowledge’ or ‘democracy’ or ‘morality’, for much the same reasons. But our aims in philosophy are different than our aims in running a political campaign. In short, they have more to do with gaining a better understanding of reality than they do with winning political office. (Or at least they *should* have more to do with that). So some of the costs of paraphrase that I outlined above are mitigated in a philosophical context. And some of the benefits of paraphrase—e.g. helping us get clearer on what exactly we are arguing about—are going to be more important in the context of doing philosophy than they are in other contexts (e.g. the context of running for political office). Therefore, when engaged in a metalinguistic negotiation, it will often be helpful to paraphrase what is going on, and thereby shift a metalinguistic negotiation to a canonical dispute about the issue in conceptual ethics that the negotiation was about.

Indeed, it will then often be helpful to refocus the dispute even more. Recall my earlier discussion of the distinction between (a) the *immediate* topic of disagreement in a metalinguistic negotiation vs. (b) the background reasons we have for engaging in that metalinguistic negotiation in the first place. For example: in the Secretariat case, I claimed that part of what drives the argu-

⁶⁰For more on this point, see Plunkett and Sundell, ‘Antipositivist Arguments from Legal Thought and Talk’.

ment between the speakers about how to use the word ‘athlete’—and part of what makes the dispute a substantive one worth having—are background issues about what kinds of animals get important kinds of social recognition and praise. In turn, the different views that the speakers have on that subject might reflect deeper disagreements they have: e.g. disagreements about the value or rights of non-human animals. So one thing that will be helpful to do in the philosophical context is this: try to figure out what those background disagreements are, and switch to having a canonical dispute that is directly about those issues. In order to figure out what those issues are, one helpful move will be this: both speakers in the dispute should try to figure out if it possible to state what their disagreement is in a way that does *not* use the term(s) under negotiation (e.g. ‘athlete’ in the Secretariat case).⁶¹

In some cases, when speakers do so, it is possible that speakers might discover that there is nothing that they disagree about *other* than an issue about concept/word pairing that was the immediate topic of disagreement in the metalinguistic negotiation. But, in general, we have good reasons to expect that will not to be the case, if the dispute really was a metalinguistic negotiation to begin with (and not, for example, a mere ‘talking past’). That is because, in general, speakers don’t just have normative views about concept/word pairing that hang totally free from other normative views they have about issues beyond conceptual ethics. Rather, those normative views are standardly connected to a whole host of other normative views. (For example, in the Secretariat case, it is hard to see how each speaker would really have different normative views about how the term ‘athlete’ should be used in that context if they didn’t *also* have different normative views about other topics, e.g. which kinds of creatures deserve praise and reward.) So, if two speakers disagree about a normative issue about concept/word pairing, we should generally expect that they do so in part because of other normative issues that they disagree about.

In moving to a dispute about these background normative issues, one thing that will be important to ask is this: whether they are about important issues that are worth arguing about. In some cases, they will be. In other cases, they won’t. Just because an issue is a *normative* one doesn’t mean that it is a

⁶¹The methodological remarks made above are closely related to claims made by Chalmers in ‘Verbal Disputes’, in discussion of his proposed ‘method of elimination’. Indeed, more generally, there are many important similarities between much of what I have argued in this paper and what Chalmers argues in that paper, as I have flagged in footnotes throughout this paper. I am much indebted to Chalmers’s work in helping me think through many of the core issues I discuss in this paper about meta-philosophy and philosophical methodology. There are also, however, some important differences between what I argue in this paper and what Chalmers argues in ‘Verbal Disputes’. In broad terms, these differences have to do with (a) the details of my explanation of how philosophers employ metalinguistic usage to communicate genuine, substantive disagreements and (b) the assessment of when a given philosophical dispute is a substantive one worth having. A full discussion of the similarities and differences between our views is, however, beyond the scope of this paper.

particularly important issue, or that it is one that it is worth spending our time arguing about.

This general point applies to normative issues in conceptual ethics. In short: some issues in conceptual ethics are important ones worth arguing about, but that doesn't mean that all of them are. In certain cases, for example, it might really matter which of a range of concepts we use a philosophical term to express moving ahead. In other cases, a disagreement about that might not matter that much at all. The issue about concept/word pairing might not matter that much. Whether it does or not will depend on a whole host of factors, including, for example, what kind of impact the choice about concept/word pairing will have on other parts of our inquiry, or other parts of our lives. Some terms in philosophy have rich connotations that are hard to shake, largely independent of the specific meaning of the term that we settle on. For example, it is hard to shake the idea that, at least when used in a philosophical context, 'knowledge' picks out something that is important for figuring out normative issues about what beliefs to have. Or, from a different angle, notice that some terms used in philosophy, e.g. 'democracy', 'freedom', 'race', or 'gender', have strong ties to pressing social/political issues outside of philosophy, and our use of them in philosophy might have an impact on how discussions go outside of philosophy. In such cases, even if the only thing that two speakers can figure out that they disagree about is how such a term should be used, that can still be a substantive disagreement well worth having. But that might not be the case for any number of other terms in philosophy.

IV. Conclusion: the importance of conceptual ethics for philosophy

The final point that I want to make concerns the centrality of conceptual ethics to work in philosophy. The framework that I have sketched in this paper (and argued for why we should take it seriously) is one on which an important dimension to many philosophical disputes is this: which of a range of competing concepts should be using for the purposes at hand? When introduced on its own, this kind of normative question can seem like a somewhat marginal one compared to more familiar normative questions in epistemology, ethics, or political philosophy. But, if the kind of framework I have suggested is onto something important about many philosophical disputes, then we should be wary of this attitude. Issues in conceptual ethics might very well be at the heart of many philosophical disputes, whether we recognize that or not.⁶² Reflection on core issues in the foundations of conceptual ethics—for example, questions such as 'what, in general, can we say about what concepts we should employ?' or 'What settles normative issues about concept-word pairing?'—might thus not only be interesting normative questions to consider in their own right, but

⁶²For more on this point, see Burgess and Plunkett, 'Conceptual Ethics I'; and Burgess and Plunkett, 'Conceptual Ethics II'. See also Chalmers, 'Verbal Disputes'.

ones central to our activity as philosophers. We would thus do well to take them seriously.⁶³

Acknowledgements

Thanks to Derek Ball, Chris Barker, Gunnar Björnsson, David Braddon-Mitchell, Mark Budolfson, Alexis Burgess, Elisabeth Camp, Herman Cappelen, David Chalmers, Matthew Chrisman, Shamik Dasgupta, Andy Egan, Stephen Finlay, Johann Frick, Lyndal Grant, Mark Greenberg, Sally Haslanger, Louis-Philippe Hodgson, Torfinn Huvenes, Frank Jackson, Sam Levey, Dustin Locke, Errol Lord, Peter Ludlow, David Manley, Berislav Marušić, Eliot Michaelson, Tristram McPherson, Hille Paakkunainen, Philip Pettit, Peter Railton, Mark Richard, Michael Smith, Justin Snedegar, Tim Sundell, Rachel Sterken, Sharon Street, Folke Tersman, Kevin Toh, Pekka Väyrynen, and Alec Walen. Versions of this paper were presented at Arché research centre at the University of St. Andrews, the Center for Study of Mind in Nature at the University of Oslo, a workshop on Normativity and Reasoning at NYU Abu Dhabi, a workshop on conceptual engineering at MIT, and at a workshop in honor of Frank Jackson in Bundanoon, Australia. Thanks to everyone who participated in those sessions for their helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Barker, Chris. 'The Dynamics of Vagueness'. *Linguistics and Philosophy* 25 (2002): 1–36.
- Barker, Chris. 'Negotiating Taste'. *Inquiry* 56, no. 2–3 (2013): 240–57.
- Braddon-Mitchell, David. 'Qualia and Analytical Conditionals'. *Journal of Philosophy* 100, no. 3 (2003): 111–35.
- Brandom, Robert. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press, 1994.
- Burgess, Alexis, and David Plunkett. 'Conceptual Ethics I'. *Philosophy Compass* 8, no. 12 (2013): 1091–101.
- Burgess, Alexis, and David Plunkett. 'Conceptual Ethics II'. *Philosophy Compass* 8, no. 12 (2013): 1102–10.
- Cappelen, Herman, and John Hawthorne. *Relativism and Monadic Truth*. Oxford: Oxford University Press, 2009.
- Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.
- Chalmers, David J. 'Two-Dimensional Semantics'. In *Oxford Handbook of the Philosophy of Language*, ed. Ernie Lepore and Barry Smith, 574–606. Oxford: Oxford University Press, 2006.

⁶³This upshot of the kind of metalinguistic view that I have advocated for in this paper is also briefly discussed at the end of Burgess and Plunkett, 'Conceptual Ethics II'.

- Chalmers, David J. 'Verbal Disputes'. *Philosophical Review* 120, no. 4 (2011): 515–66.
- Chalmers, David J., and Frank Jackson. 'Conceptual Analysis and Reductive Explanation'. *Philosophical Review* 110 (2001): 315–60.
- Copp, David. 'Moral Naturalism and Three Grades of Normativity'. In *Normativity and Naturalism*, ed. Peter Schaber, 7–45. Frankfurt: Ontos-Verlag, 2005.
- Dworkin, Ronald. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press, 2011.
- Egan, Andy. 'Epistemic Modals, Relativism, and Assertion'. *Philosophical Studies* 133, no. 1 (2006): 1–22.
- Enoch, David. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2011.
- Fine, Kit. 'Guide to Ground.' In *Metaphysical Grounding: Understanding the Structure of Reality*, ed. F. Correia and B. Schnieder, 37–80. Cambridge: Cambridge University Press, 2012.
- Gibbard, Allan. *Thinking How to Live*. Cambridge, MA: Harvard University Press, 2003.
- Gibbard, Allan. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press, 1990.
- Haslanger, Sally. 'Gender and Race: (What) Are They? (What) Do We Want Them to Be?' *Nous* 34, no. 1 (2000): 31–55.
- Haslanger, Sally. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press, 2012.
- Horn, Laurence R. *A Natural History of Negation*. Chicago, IL: University of Chicago Press, 1989.
- Jackson, Frank. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon, 1998.
- Kaplan, David. 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives.' In *Themes from Kaplan*, ed. Joseph Almog and John Perry, 481–563. Oxford: Oxford University Press, 1989.
- Kennedy, Chris. 'Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Predicates'. *Linguistics and Philosophy* 30, no. 1 (2007): 1–45.
- Lewis, David. 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy* 61, no. 4 (1983): 343–77.
- Lewis, David. 'Scorekeeping in a Language Game'. *Journal of Philosophical Logic* 8 (1979): 339–59.
- Ludlow, Peter. 'Cheap Contextualism'. *Philosophical Issues* 18 (2008): 104–29.
- Ludlow, Peter. *Living Words: Meaning Undetermination and the Dynamic Lexicon*. Oxford: Oxford University Press, 2014.
- MacFarlane, John. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press, 2014.
- Margolis, Eric, and Stephen Laurence. *Concepts: Core Readings*. Cambridge, MA: MIT Press, 1999.
- Marques, Teresa, and Manuel García-Carpintero. 'Disagreement about Taste: Commonality Presuppositions and Coordination'. *Australasian Journal of Philosophy* 92, no. 4 (2014): 701–23.
- McKenna, Michael and Coates D. Justin. 'Compatibilism.' In *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), ed. Edward N. Zalta. <http://plato.stanford.edu/archives/sum2015/entries/compatibilism/>.
- McLaughlin, Brian, and Bennett Karen. 'Supervenience.' In *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), ed. Edward N. Zalta. <http://plato.stanford.edu/archives/spr2014/entries/supervenience/>
- McPherson, Tristram. 'Against Quietist Normative Realism'. *Philosophical Studies* 154, no. 2 (2011): 223–40.
- Parfit, Derek. *On What Matters*. 2 vols. Oxford: Oxford University Press, 2011.
- Plunkett, David, and Timothy Sundell. 'Antipositivist Arguments from Legal Thought and Talk: The Metalinguistic Response'. In *Pragmatism, Law, and Language*, ed. Graham Hubb and Douglas Lind, 56–75. New York: Routledge, 2014.

- Plunkett, David, and Timothy Sundell. 'Disagreement and the Semantics of Normative and Evaluative Terms.' *Philosophers' Imprint* 13, no. 23 (2013): 1–37.
- Plunkett, David, and Timothy Sundell. 'Dworkin's Interpretivism and the Pragmatics of Legal Disputes.' *Legal Theory* 19, no. 3 (2013): 242–81.
- Putnam, Hilary. *Reason, Truth, and History*. Cambridge: Cambridge University Press, 1981.
- Railton, Peter. 'Moral Realism'. *The Philosophical Review* 95 (1986): 163–207.
- Rosen, Gideon. 'Metaphysical Dependence: Grounding and Reduction'. In *Modality: Metaphysics, Logic, and Epistemology*, ed. Bob Hale and Aviv Hoffmann, 109–36. Oxford: Oxford University Press, 2010.
- Rosen, Gideon. 'Real Definition.' *Analytic Philosophy* (Forthcoming).
- Schroeter, Laura, and Francois Schroeter. 'Normative Concepts: A Connectedness Model.' *Philosophers' Imprint* 14, no. 25 (2014): 1–26.
- Sider, Theodore. *Writing the Book of the World*. Oxford: Oxford University Press, 2012.
- Stalnaker, Robert. 'Assertion.' In *Context and Content: Essays on Intentionality in Speech and Thought*, 78–95. Oxford: Oxford University Press, 1999.
- Stevenson, Charles L. 'Persuasive Definitions'. *Mind* 47, no. 187 (1938): 331–50.
- Sundell, Tim. 'Disagreement, Error, and an Alternative to Reference Magnetism'. *Australasian Journal of Philosophy* 90, no. 4 (2012): 743–59.
- Sundell Tim. "Disagreements About Taste". *Philosophical Studies* 155, no. 2 (2011): 267–88.
- Sundell, Tim. 'Eligibility and Ideology in the Vat.' In *The Brain in a Vat*, ed. Sanford Goldberg. Cambridge: Cambridge University Press, Forthcoming.
- Thomasson, Amie L. 'Answerable and Unanswerable Questions.' In *Metametaphysics: New Essays on the Foundations of Ontology*, ed. David J. Chalmers, David Manley, and Ryan Wasserman, 444–71. Oxford: Oxford University Press, 2009.
- Trogon Kelly. 'An Introduction to Grounding'. In *Varieties of Dependence*, ed. M. Hoeltje, B. Schnieder, and A. Steinberg, 97–122. Munich: Philosophia Verlag, 2013.
- Wedgwood, Ralph. *The Nature of Normativity*. Oxford: Oxford University Press, 2007.