



**Inquiry**

An Interdisciplinary Journal of Philosophy

ISSN: 0020-174X (Print) 1502-3923 (Online) Journal homepage: <https://www.tandfonline.com/loi/sinq20>

## Metalinguistic negotiation and speaker error

David Plunkett & Tim Sundell

To cite this article: David Plunkett & Tim Sundell (2019): Metalinguistic negotiation and speaker error, *Inquiry*, DOI: [10.1080/0020174X.2019.1610055](https://doi.org/10.1080/0020174X.2019.1610055)

To link to this article: <https://doi.org/10.1080/0020174X.2019.1610055>



Published online: 04 Jul 2019.



Submit your article to this journal [↗](#)



Article views: 64



View related articles [↗](#)



View Crossmark data [↗](#)



# Metalinguistic negotiation and speaker error

David Plunkett<sup>a</sup> and Tim Sundell<sup>b</sup>

<sup>a</sup>Department of Philosophy, Dartmouth College, Hanover, NH, USA; <sup>b</sup>Department of Philosophy, University of Kentucky, Lexington, KY, USA

## ABSTRACT



In recent work, we have argued that a number of disputes of interest to philosophers – including some disputes amongst philosophers themselves – are *metalinguistic negotiations*. Prima facie, many of these disputes seem to concern worldly, non-linguistic issues directly. However, on our view, they in fact concern, in the first instance, normative questions about the use of linguistic expressions. This will strike many ordinary speakers as counterintuitive. In many of the disputes that we analyze as metalinguistic negotiations, speakers might quite strongly resist the idea that their debate is in any sense about language. In this paper, we explore and provide responses to what we take to be the best versions of an objection that our view involves an unacceptable attribution of false beliefs to ordinary speakers.

**ARTICLE HISTORY** Received 9 January 2019; Accepted 15 January 2019

**KEYWORDS** Metalinguistic negotiation; speaker error; conceptual ethics; disagreement; conceptual engineering

## 1. Introduction

In recent work, we have argued that a number of disputes of interest to philosophers – including some disputes amongst philosophers themselves – are *metalinguistic negotiations*. Metalinguistic negotiations involve a disagreement between speakers about how they should use language in the context at hand. In a metalinguistic negotiation, speakers navigate such normative issues about words and concepts *implicitly*, via a “metalinguistic” use of a term, wherein speakers use words to communicate views about the very term they are using. In our work thus far, we have argued that metalinguistic negotiations occur in ethics, epistemology, aesthetics, metaphysics, law, and politics, and we have explored some of the implications for these domains.<sup>1</sup>

**CONTACT** David Plunkett  david.plunkett@dartmouth.edu  Department of Philosophy, Dartmouth College, 6035 Thornton Hall, 19 College Street, Hanover, NH 03755, USA

<sup>1</sup>The relevant joint work is (Plunkett and Sundell 2013a), (Plunkett and Sundell 2013b), and (Plunkett and Sundell 2014). The relevant single-authored work from Plunkett is (Plunkett 2015) and (Plunkett 2016).

An important distinction among metalinguistic disputes is how aware the speakers are of what they are doing. In some cases, they might understand that they are engaged in a metalinguistic dispute. In other cases, they might not. Indeed, they might resist this characterization quite strongly. One important objection to the idea of metalinguistic negotiation is based on this type of resistance. To see the force of such an objection, consider some central motivations behind a number of competing theories.

Laura and François Schroeter take speaker intuitions of sameness of reference as the starting point for their “connectedness model”. Speakers seem to take themselves to mean the same thing by their words, no matter how divergent their beliefs about the object named. This observation grounds Schroeter and Schroeter’s notion of *de jure* sameness of reference – sameness of reference written into the rules of the conversational game – and that notion is at the core of their model (Schroeter and Schroeter 2014).

Derek Ball motivates his temporal externalism by looking to ordinary intuitions about continuity of content across sometimes major theoretical change. Because we take ourselves to be talking about the same thing as future speakers, what we’re talking about must be in some way sensitive to considerations that will only become apparent later in time. Ball’s theory aims to elaborate what that kind of sensitivity would need to be like in order to bear out our intuitions of conceptual continuity (Ball, forthcoming).

Herman Cappelen criticizes our work on metalinguistic negotiation by drawing attention to how ordinary speakers engaged in the relevant disputes would seem to find explicitly linguistic moves in the debate to be off topic. Ordinary speakers would, for example, treat proposals that we change the meaning of the term in question and stipulate a new definition for it to be irrelevant to the worldly issues they take themselves to be debating. Similarly, they would take proposals about changing the spelling of the term in question to be irrelevant. Speakers, in other words, seem not to understand their own debates as being about language in any sense. Cappelen claims that his own theory of these cases avoids this concern (Cappelen 2018, 174–175).

What all of these theories have in common is that they aim to hew closely to speakers’ intuitive grasp of their own utterances and conversations. Each theory, in its own ways, explores what meanings would have to be like if they were the kind of things ordinary speakers seem to take them to be. At

---

The relevant single-authored work from Sundell is (Sundell 2011), (Sundell 2012), (Sundell 2016b), and (Sundell 2016a).

least *prima facie*, this is in stark contrast to our metalinguistic account, which acknowledges that ordinary speakers might, in certain contexts, find a metalinguistic analysis of their debate to be highly counterintuitive. This distance between ordinary speaker self-understanding and our theoretical analysis has the potential to form the basis of a serious objection to the view. After all, if hewing closely to speakers' self-understanding of their utterances and conversations is a point in favor of these other views (as is suggested by the way their proponents argue for them), then surely it is a theoretical cost to our view that it does not.<sup>2</sup>

In this paper, we consider what we take to be the best versions of an objection that our view involves an unacceptable attribution of false beliefs to ordinary speakers. In previous work, we have briefly responded to this type of objection, in the context of responding to a prominent argument for antipositivism in legal philosophy (Plunkett and Sundell 2014). In this paper, we expand on that response, present it in more general terms, and then develop additional lines of response. We grant that the attribution of error can in some circumstances be a theoretical cost. However, it is important to get clear on *why* and *when* this would be a cost. Once we do, we claim the following: (a) it's not obvious that speakers have the kinds of theoretically sophisticated beliefs which would, on our metalinguistic view, be in error, and (b) even if they do, our metalinguistic framework *reduces* the theoretically problematic kind of error (speakers unaccountably misusing their own words), while the error it potentially attributes (speakers holding to mistaken folk-linguistic theories) is less worrisome.

## 2. Metalinguistic negotiation

Our view of metalinguistic negotiation can be glossed as follows. Speakers involved in a metalinguistic negotiation disagree about how they should use language in their context. Sometimes speakers argue about these

---

<sup>2</sup>Not every analysis of the kinds of dispute at focus here are deferential to speaker intuition in the ways that the views from Schroeter and Schroeter, Ball, and Cappelen are. For example, the views advocated for in (Björnsson and Finlay 2010), (Finlay 2016), (Khoo and Knobe 2016), and (Silk 2016) have certain elements in common with our views on metalinguistic negotiation that might open them up to a similar type of worry. In particular, on these other views, a number of disputes will be analyzed as non-canonical in ways that many speakers might resist (even if they are not analyzed as "metalinguistic" in our sense). Elements of our response to this kind of objection will aid these views as well, though there are of course different resources that each of these views can bring to bear. It is beyond the scope of this paper to go into detail about the relevant differences here. We take it that – in part because of our insistence on the "metalinguistic" nature of these negotiations – our view is open to a particularly strong version of this objection. Our hope is that, by working through how this objection goes with respect to our own view, we can lay the foundations for a more thorough examination of how this objection fares with respect to views in this neighborhood.

issues *explicitly*, in terms of the literal content of what they assert. But, in metalinguistic negotiations, these disagreements are expressed tacitly. Such disputes do not necessarily involve speakers “merely talking past” one another. To the contrary, these disputes often express genuine, substantive disagreements that can be well worth having. Our choices about which concepts to use, and about which concepts to pair with which words, can be loaded with significance, in ways that correspond closely to the object-level matters that intuitively are at issue in the disputes in question.<sup>3</sup>

Let’s unpack this a bit.

### 2.1. *Disagreements vs. disputes*

We distinguish between *disagreement* on the one hand, and *disputes*, on the other. *Disagreement*, as we use the term, indicates a kind of rational conflict in mental states. This notion of disagreement takes it to be a *state* – roughly, two speakers accepting conflicting contents  $p$  and  $q$  – rather than an *activity*. It is neutral with respect to the nature of the conflicting contents. ( $P$  could be a proposition that entails not- $q$ , or  $p$  could be a preference or plan, the satisfaction or completion of which is incompatible with  $q$ , etc.) Finally, it divorces the notion of disagreement from any particular interaction in which the disagreement is expressed. People who have never met and will never meet can count as disagreeing with one another.

*Disputes*, on the other hand, are *linguistic exchanges*. In particular, disputes are linguistic exchanges that appear to evince or express a disagreement. Appear to whom? They might appear that way to an observer, to the participants themselves, or to a theorist considering the dispute. Some disputes may succeed in expressing genuine disagreements, while other disputes may not. And some disputes may succeed in expressing genuine disagreements that are worth our time arguing about in a linguistic exchange, while others might express merely trivial disagreements that aren’t worth arguing over.

### 2.2. *Canonical vs. non-canonical disputes*

Some disputes express disagreement over the literally expressed semantic content of the expressions used in the dispute.<sup>4</sup> In this kind of dispute, the

<sup>3</sup>We have defended this basic framework, and many of the details below, in previous work. See footnote 1 for the relevant references.

<sup>4</sup>We set aside here arguments made by (Yalcin 2014), (Ninan 2010), and (Rabern 2012) that notions like “semantic” or “compositional value” should be kept carefully distinct from notions like “assetoric

content that the speakers *say* (in the technical sense of “what is said”) is the very content about which they disagree. In deference to the widespread assumption that this is the most common or typical way to express a disagreement, we call disputes of this kind *canonical disputes*.

In other disputes, the speakers do express a disagreement, but the content about which they disagree isn’t the semantic content of their expressions in their context. The content about which they disagree might instead be conversationally implicated, or presupposed, or conventionally implicated, etc. When we speak, we communicate a broad range of information, and much of that information is available as something for our listener to express a disagreement with. When the dispute expresses a disagreement about content other than what is said in the dispute, we call the dispute “non-canonical”.

### 2.3. *Metalinguistic usage*

*Metalinguistic disputes* are one type of non-canonical dispute. To understand the notion of *metalinguistic dispute*, we have to begin with the notion of a *metalinguistic usage*.

A *metalinguistic usage* of a term is a case where that term appears to be used (not mentioned) to convey information about how that very term is or ought to be used in the context. In Barker (2002)’s much-discussed example, someone asks you what counts as “tall” around here, and you reply, “Well, Feynman is tall” as you and your listener both look over at Feynman (Barker 2002). We could even imagine that Feynman just happens to be standing in front of a measuring stick.

In this case, the central communicative upshot of your utterance is not new information about Feynman’s height, which, after all, is mutually known by you and your listener. Rather, you’ve communicated information about language, and in particular, about the local height threshold for “tall”.

### 2.4. *Metalinguistic disputes*

When two speakers employ competing metalinguistic usages of a term to express a disagreement about how that term is used in the context, or how it ought to be used, we call that a *metalinguistic dispute*. If one speaker uses

---

content”, “proposition expressed”, or “what is said”. We suspect that these arguments will prove helpful for our overall project, but the points are subtle, and beyond the scope of this paper.

a term metalinguistically to communicate either descriptive or normative information about that term's use, another speaker is entirely free to use that same term metalinguistically to communicate a conflicting view. They might, for example, reply to you, in the context of Barker's scenario, that "No, Feynman isn't tall".

### **2.5. Descriptive vs. normative metalinguistic disputes**

When a metalinguistic dispute centers on the question of how a term is used in the context, this is a *descriptive metalinguistic dispute*.

When a metalinguistic dispute centers on the question of how a term *ought* to be used in the context (or, relatedly, on *how to* use the term in the context), this is a *normative metalinguistic dispute*. Normative metalinguistic disputes concern practical issues about how to proceed using the term, or normative/evaluative assessment of the various options for doing so. *Metalinguistic negotiation* is another word for *normative metalinguistic dispute*.

### **2.6. Metalinguistic disputes about character**

Some metalinguistic disputes (descriptive or normative) concern the resolution of context-sensitivity for a context-sensitive linguistic expression. For example: speaker A thinks the threshold for counting as 'tall' is 6 feet, while speaker B thinks it is 6.5 feet. In this case, the speakers agree on the basic context-insensitive meaning of the term 'tall'.

In other cases, this is not true. Consider Peter Ludlow's Secretariat case, which we discuss at length in other work.<sup>5</sup> In that case, two speakers are arguing on a sports radio show about a list of the 50 greatest athletes of the 20th-century recently published by Sports Illustrated. The racehorse Secretariat was on that list. Speaker A says that "Secretariat was an athlete" and speaker B says that "No, Secretariat was not an athlete". They agree on all the seemingly relevant empirical facts: e.g. how many races Secretariat won, how fast he was, etc. But speaker B, it turns out, is disposed never to use the term 'athlete' to refer to any non-human animals (even very successful race horses such as Secretariat).

We claim, drawing on Ludlow, that these speakers are advancing different normative views about what the term 'athlete' should mean.

---

<sup>5</sup>This case is drawn from (Ludlow 2008). For our discussion of it, see (Plunkett and Sundell 2013a) and (Plunkett and Sundell 2013b).

Their proposals do not involve different claims about how to set a threshold for piece of context-sensitive terminology. Nor do they concern how to make precise a vague term. Rather, we claim, they involve rival proposals for the meaning of the term ‘athlete’: different proposals about what the character of the term should be. If we take the meanings of words to be the concepts they express, we can put this point another way. The speakers are proposing rival views about which concept the term should be paired with, and thus which concept should be deployed in the context at hand. This kind of issue about which concepts we should use is an issue in *conceptual ethics*.<sup>6</sup>

## 2.7. Metalinguistic negotiation and object-level issues

Ludlow’s Secretariat case helps us bring out another important point. This is that metalinguistic negotiations can be a way of arguing about object-level issues (issues not about thought and talk).

Think of some of what is at stake in whether Secretariat is called an ‘athlete’ or not. This might include who gets certain kinds of fame, praise, or recognition. For an animal rights activist, having Secretariat on the list might well change how people view animals. That might convince someone that there are practical benefits to calling non-human animals ‘athletes’, given the resonance of the term ‘athlete’ in our culture and given the broad functional role that term plays in our discussions.<sup>7</sup> On the other hand, one might think there are particular kinds of virtues that only humans (among existing animals) can exemplify, and it is best to use the term ‘athlete’ to help us keep track of those particular virtues. This suggests that it might well be an important normative topic how we should use the term ‘athlete’ in this context. And if it is an important topic, it remains so whether we are argue about it *explicitly* (in a canonical dispute) or *more indirectly* (for instance, via metalinguistic negotiation).

To bring this point out, we have also used the example of conflicting normative views about how to define the word ‘torture’. Consider the following definitions:

---

<sup>6</sup>We here use the term ‘conceptual ethics’ to cover normative and evaluative questions about concepts, including, for example, questions about which concepts a given agent A should use in context C, and questions about which concepts are better or worse. See (Burgess and Plunkett 2013a) and (Burgess and Plunkett 2013b). Some philosophers prefer to use the term ‘conceptual engineering’ to cover these same kinds of issues. See, for example, (Scharp 2013), (Eklund 2017), and (Cappelen 2018).

<sup>7</sup>We do not present a theory of “resonance” here or “functional role” of terms. But there are connections here to Sally Haslanger’s work, which emphasizes similar points about the resonance of key terms. See (Haslanger 2000). See also (Thomasson, forthcoming). (What we discuss is similar to the notion of ‘lexical effects’ as discussed in [Cappelen 2018]).



UN (1984): 'torture' = any act inflicting severe suffering, physical or mental, in order to obtain information or to punish

US dept. of justice (2002): 'torture' = any such act inflicting pain rising to the level of death, organ failure, or the permanent impairment of a significant body function.

Should we be neutral as to which definition we use in legal documents in the USA? What about in political discourse in the USA? What about in discussion amongst moral and political philosophers? We suggest that, regardless of what turns out to be analytic of the term 'torture', that term has important resonance in our current social-historical context. It can matter which definition we use for the kinds of ways people think of certain acts by default, and what normative/evaluative inferences they are prone to make. Based on this, it can matter, in many contexts, how we choose to use the word 'torture'. (How much it matters is a separate question, and something that can vary between different contexts).

The basic idea is this. People who have different views on how we should use the term 'torture' are likely to have these views *because of* further moral/social/political/legal views they have – views involving object-level issues about how to live, and what to do, rather than about words and concepts. Within a metalinguistic negotiation over the term 'torture', these disagreements about *object-level* matters are things they are able to express *via* expressing a disagreement about how to use the word 'torture'. The *immediate* disagreement expressed is one in conceptual ethics, about how we should use words or concepts. But people's views in conceptual ethics are almost always based in views they have about things *other* than *representational-level* issues about words or concepts. Those are often the deeper disagreements in effect driving the metalinguistic negotiation, and are about the issues that ultimately matter more.

Similar points can be made about terms involving *descriptive* issues, rather than normative ones. For example: think of debates about how the term 'species' should be used in the context of doing biology. Or about the term 'semantics' in linguistics. Or the term 'market' in economics. How these representation-level issues get resolved makes a difference for how we proceed in our inquiries, just as they make a difference in our lives more generally. The idea here is nicely summed up in the following passage from David Chalmers:

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. (Chalmers 2011, 542)

In other words, for creatures like us, how we use words matters.

### 3. Speaker error

An important distinction among metalinguistic disputes is how aware the speakers are of what they are doing in the dispute. In some cases, they might understand that they are engaged in a metalinguistic dispute. In other cases, they might not. Indeed, they might resist this characterization quite strongly. One important objection to the idea of metalinguistic negotiation is based on this type of resistance.

Here's how such an objection might go. Metalinguistic negotiations express disagreements that, in the first instance, are about language and thought—about the resolution of context sensitivity, or the precisification of vague terms, or which concept to pair with some term with a range of candidate meanings, and so on. That means that to the extent that these disputes reflect disagreements about non-linguistic *object-level* matters they do so at best indirectly.

For example, in our discussions of disputes about whether waterboarding is “torture”, we claim that the speakers disagree, in the first instance, about how to use the word ‘torture’ in their context – about which of two competing definitions of that word should be used. In these cases, we say that each speaker literally expresses a true proposition, given what they mean by the word ‘torture’.<sup>8</sup> But even if we grant that they both speak truly in terms of the literal content they express, these speakers also suggest that their definition is appropriate to the circumstances. That disagreement about language and thought reflects disagreements about the treatment of prisoners – or so we claim – but it does so only *via* a disagreement about how to use a word.

The problem, according to the objection, is that debates about torture don't *seem* to be about language *at all*. They seem to concern matters about the treatment of prisoners, and they seem to do so *directly*, not in some indirect, roundabout way. If you ask participants in these disputes what their argument is about, they will say that it is about torture, not about the word ‘torture’, about a controversial practice, not about our choice of concepts.<sup>9</sup>

---

<sup>8</sup>This reflects the individualist conception of meaning that we favor (at least for many explanatory projects involving thought and talk). This is an assumption not all proponents of the basic idea of metalinguistic negotiation need accept. For example, see (Thomasson 2016).

<sup>9</sup>For an example of this kind of objection to our view, see (Cappelen 2018, 174–175).

To analyze such a dispute as a metalinguistic negotiation is thus, according to the objection, to attribute to the participants a set of false beliefs about their own activity. In particular, in believing that their dispute is an ordinary, canonical dispute concerning the object-level issues that intuitively are at issue, speakers are wrong in two respects. First, they are mistaken in believing that the claims over which they disagree are the literal semantic content of their utterances. (In fact, the contents they literally express might not even conflict.) Second, they are mistaken in believing that their disagreement is not, in the first instance, about language and thought. (In fact, it is.) Attributing widespread false beliefs to ordinary speakers is, potentially, a theoretical cost. Since we mean for our analysis to apply not just to disputes among “ordinary” speakers, but also to lawyers, judges, policy makers, and philosophers – many of whom are highly attuned to language – this cost may be even higher than usual. We’ll call this kind of worry a “speaker error” objection.

We think that responding to speaker error objections requires getting clear on why a theory that attributes mistaken beliefs is thought to bear some extra explanatory burden, and on what, exactly, that burden is. What, exactly, is wrong with attributing false linguistic beliefs to users of a language? Maybe the mistake is in thinking that speakers are likely to have *true* beliefs! After all, in the history and philosophy of science, it’s often claimed that scientists can be quite bad at understanding what they are doing when they are engaged in scientific activity. Perhaps there is no reason to think that the situation is any different in the case of law, politics, or philosophy. Why should we expect practitioners to be adept sociologists and anthropologists of their own activity?

We are sympathetic to this line of thought, and it has some role to play in our response. Practitioners in any domain of human practice are often bad theorists about what activity in that domain involves. However, this is not the end of the story. For even if this thought is correct, it would still be good to offer an explanation of why that error occurs – an explanation that goes beyond just saying “practitioners are bad theorists”. There are two ideas that motivate this thought.

First, it would be good to have an explanation of why the practitioners are prone to making the *particular* types of error they do. There are many ways for someone to be mistaken about what they are doing. Insofar as a theory ascribes a specific type of mistake to the relevant speakers or agents, it owes some accounting of why that mistake is the one that is made rather than others (see Silk 2016, for more on this point). This theoretical requirement holds, no matter

how unsurprising it may be, in general, that the speakers or agents will be mistaken in some way or other.

Second, other things being equal, a linguistic theory in particular should avoid positing widespread error to ordinary speakers. Why? In fields where speakers' linguistic judgments and usage are among the primary data for theorizing, blanket attributions of error create a worrisome distance between the theory and the facts to which that theory is meant to be accountable. If we feel too free to describe speakers as wrong, our theory will too easily write off potentially disconfirming data as noise. Semantic theories are of course not required to hew perfectly to speaker usage and intuition, any more than theories of syntax or phonology are. Everyone agrees that there is some role for a notion of "performance error". But if you want to write off some inconvenient speaker intuition as a mistake, you should be able to offer some independently motivated account of where that mistake is coming from (see Sundell 2011 for discussion). Otherwise, the data can start to look evidentially irrelevant to the theory. Thus, whether a theorist is sympathetic to widespread error in general, or on the contrary, considers the positing of systematic error to be a theoretical cost, one should, either way, have some explanation to offer of the specific types of error she posits.

That means that the relevant questions for us are: (1) How much, and what kind of error are we in fact committed to attributing to speakers? and (2) Do we have something independently plausible to say about why speakers make mistakes of that kind?

As suggested above, there are potentially two kinds of ordinary beliefs that, on our metalinguistic view, are in error. First, speakers may be mistaken in believing that their dispute is canonical. Thinking of "canonical disputes" in terms of propositions, we can put the point as follows: speakers may believe that their utterances semantically express conflicting propositions, as opposed to semantically expressing *consistent* propositions the expression of which, in context, communicates certain further claims that do manage to conflict. Second, relatedly, speakers are mistaken in thinking that their dispute expresses a disagreement that directly concerns the worldly, first-order matters that intuitively are at issue. This is opposed to them concerning, in the first instance, matters of language and thought that, given certain background assumptions, are closely tied to (and perhaps run largely in parallel with) the worldly, first-order matters that intuitively are at issue. We treat these issues in turn.

### 3.1. False beliefs about linguistic mechanism

With respect to speaker beliefs about how – semantically or pragmatically – their conflicting views are expressed, the force of this worry will of course depend on what the mechanism or mechanisms for metalinguistic usage turn out to be. This is an issue on which we have been content to remain neutral. If, for example, a Barker-style dynamic analysis of metalinguistic usage wins the day, then there's no clear sense in which the metalinguistic update effects of a certain usage are less *semantic* than the effects on those elements of the common ground that involve first-order, non-linguistic matters. On Barker's picture, contents are evaluated at world/discourse pairs. The context set tracks possibilities, uncertainty, and indeterminacy concerning both what the non-linguistic world is like and concerning the state of the discourse. A successful assertion serves to eliminate possibilities of both kinds from the context set.<sup>10</sup> If this kind of analysis is right, then there's no clear sense in which the proponent of the metalinguistic approach is required to describe as mistaken speakers who take the main communicative upshot of their speech to be part of "what's asserted" or "what is said". The categorization of metalinguistic update effects as semantic or pragmatic is similarly obscure on theories according to which the Stalnakerian diagonal, or Perry-style "proposition created", are the main communicative upshot of the bulk of our speech.<sup>11</sup>

Other analyses, however, might prove more problematic in this regard. Suppose – as we tentatively did in some previous work – that metalinguistic usage is something more like a Grician implicature, an inference performed by the listener, over and above their interpretation of the semantic content expressed.<sup>12</sup> On such an analysis, "what the speaker said" would be the (true) proposition they express, given what they mean by the terms in question. The listener would then go on to reason – in a manner similar to how Grice describes a quality implicature – that the speaker would only have said the thing they did, using the terms in the way they did, if they (the speaker) believed that using the terms in that way was appropriate under the circumstances.<sup>13</sup> If such an analysis

<sup>10</sup>See (Barker 2013). If this turned out to be the right approach to the mechanism, then it might be that metalinguistic negotiations would fall on the other side of the *canonical/non-canonical* divide, as we currently define it. That would be an interesting development, but would not undermine the overall view.

<sup>11</sup>Thanks to Andy Egan for helpful discussion on this point.

<sup>12</sup>See (Plunkett and Sundell 2013a) and (Plunkett and Sundell 2013b). For discussion about this broadly Grician way of thinking about metalinguistic negotiation, see (Thomasson 2016).

<sup>13</sup>This would not exactly count as a quality implicature, as Grice conceives of it (Grice 1989). The maxim of quality requires that we not say things that are false, or for which we lack adequate evidence. Expressing a proposition that's true and well supported by the evidence, but that carves up the world in some inappropriate way or uses words in a way that we know to have pernicious effects is clearly uncooperative,

were right, and if ordinary speakers nevertheless believed that the content about which they disagreed was what they semantically expressed, then we would have to describe such speakers as mistaken.

Our primary response to this worry is that if there is a speaker error objection in the vicinity here, it is a worry *not* for the claim that certain usages are (in part) metalinguistic. Rather, it is a worry for specific theories of the mechanism(s) underlying metalinguistic usage. As we've emphasized, the fact that certain uses of linguistic expressions manage to communicate information about those very same expressions is there for anyone to see.<sup>14</sup> And that's all that's required to get the notion of metalinguistic negotiation up and running. If some theories of how that metalinguistic information is communicated happen to conflict with ordinary speaker intuition, then perhaps that provides *pro tanto* reason to reject those theories in favor of competing theories of how the metalinguistic information is communicated. But it does not in any obvious way provide reason for concluding that metalinguistic information is not being communicated, that such metalinguistic information could not form the basis of a disagreement, or that such a disagreement could not serve to express a closely related disagreement about object-level issues.

Even setting this aside, however, we submit that the attribution of error to ordinary speakers *still* does not succeed as an objection to metalinguistic negotiation, even supposing for the sake of argument that we are committed to the kind of properly pragmatic analysis considered directly above. We have provided some responses to this narrowly circumscribed kind of worry in earlier work (see Plunkett and Sundell 2014). We expand on them here.

Recall that this version of the mechanism worry boils down to the fact that speakers take the communicative action to be happening at the level

---

but it doesn't violate either of those stipulations. Still, the need to choose appropriate concepts and pair them with the appropriate linguistic items would seem to be in the same spirit.

<sup>14</sup>This general point goes much farther back than Barker's much-discussed Feynman example. For example, in *The Language of Morals*, Hare observes that

Sometimes we use ['good' and 'red'], not for actually conveying information, but for putting our hearer into a position subsequently to use the word "good" or "red" for giving or getting information. Suppose, for example, that he is utterly unfamiliar with motor-cars in the same sort of way as most of us are unfamiliar with horses nowadays, and knows no more about motor-cars than is necessary in order to distinguish a motor-car from a hansom cab. In that case, my saying to him "M is a good motor-car" will not give him any information about M, beyond the information that it is a motor-car. But if he is able then or subsequently to examine M, he will have learnt something. He will have learnt that some of the characteristics which M has are characteristics which make people – or at any rate me – call it a good motor-car. (Hare 1952/1963, 113)

of semantics, when really it is happening at the level of pragmatics. We have two responses to such a worry.

First, drawing distinctions between semantic and pragmatic modes of communication involves technical, theoretical notions that ordinary speakers arguably do not have at their disposal. Given the role played by theoretical notions from empirical syntax, semantics, and pragmatics in even making the relevant distinctions, it is not even clear that we should think of ordinary speakers as having the relevant concepts. Therefore, it is not clear that ordinary speakers have relevant intuitions in the first place. While it's true that some speakers may have intuitions they would describe as judgments about what was "said" rather than "implied", extreme caution is called for in projecting from ordinary judgments like these to judgments involving the corresponding theoretical notions deployed. Grice himself is careful to emphasize that his notion of "say" – which is meant to be "closely related to the conventional meaning of the words" a speaker utters – is a specialized or even artificial sense of the term (Grice 1989). It's not at all obvious why ordinary intuitions about what a speaker "says" would be tracking the same distinctions that matter to theorists working in semantics and pragmatics. Therefore, it's correspondingly unclear that those intuitions, if they are seemingly in conflict with theoretical claims that happen to be made using homophonous technical terms, are thereby in error.

Second, even if we allow that ordinary speakers have beliefs about the semantics/pragmatics distinction, that distinction can be subtle, to say the least. An ordinary speaker might *have* the relevant technical concepts but not be particularly adept at deploying them. (Compare: a college student might take a class in advanced physics and acquire some new technical concepts. But that doesn't mean she will be an expert at deploying them.) There is correspondingly only a small theoretical cost to ascribing error to those beliefs.

In the case of some pragmatic types of communication, the distinction is quite clear of course, even to ordinary speakers. For example, consider the dialogue in (1):

(1)

- (a) Care to see a movie tonight?
- (b) I've got work to do.

The speaker of (1b) has not literally said that she is not able to see a movie tonight. But that is clearly what she communicates. Should we expect the

speaker of (1b) to be aware, or easily able to become aware, that she has not literally expressed the proposition that she cannot see the movie tonight? Of course. If we were to ask the speaker, “What exactly did you say in response to the question?” we might expect an answer like “Well, all I actually said was that I have work, but obviously that means I can’t go to the movie”.

But not all cases of pragmatic communication are so easily spotted. Consider a case of relevance implicature.

(2)

(a) Who was at the party?

(b) Sarah and Danny were both able to make it.

On Grice’s account, the speaker of (2b) literally expresses only that Sarah and Danny had, in some sense, the ability to attend the party. But, on the assumption that she is being cooperative as an interlocutor, her listener can infer that Sarah and Danny not only had the ability, but in fact attended the party. As Grice emphasizes, what is important to the analysis is that such a line of reasoning *could* be undertaken by the listener: the speaker said that Sarah and Danny had the ability to attend; if they were able to attend but didn’t actually come, then the utterance would be unhelpful; but the speaker was probably trying to be helpful; so Sarah and Danny must have actually attended the party. This line of reasoning must be *available*, but it’s no part of the analysis that the listener must actually run through it. Rather, on the Gricean picture, the listener might arrive at the actual, intended meaning essentially automatically, without the need to reason on the basis of any particular assumptions about what is semantically, rather than pragmatically, communicated. And in this case, that’s a good thing. It’s actually quite strange to think of the speaker of (2b) as not having said that Sarah and Danny attended the party. It would be slippery indeed for the speaker of (2b) to argue that she wasn’t lying even though Sarah and Danny never showed up, and it’s easy to imagine it being quite difficult to persuade her listeners that she never said that Sarah and Danny were there. The distinction between literal content and relevance implicature is interesting and important to the theorist. But it’s a subtle distinction; one that ordinary speakers don’t have much reason to care about in most contexts, and would not necessarily have accurate judgments about.

Finally consider a classic case of purported quantity implicature.



(3)

(a) How many cats does Randall have?

(b) He has three cats.

According to Grice and later theories in that tradition, the speaker of (3b) has literally expressed only that Randall has at least three cats. But that of course is consistent with Randall having an additional seven cats, for a total of ten. Reasoning of the familiar sort allows the listener to infer the logically stronger proposition that Randall has exactly three cats. So the Grician story goes. But this analysis is hardly uncontroversial. Some linguists have argued that expressions like (3b) *do* literally express the *exactly* meaning (see, for example, Chierchia 2004). If these alternative theories are correct, then it is part of the literal content of (3b) that Randall has exactly three cats (rather than at least three cats), and it was wrong to think that the *exactly* part was pragmatic and not semantic. This remains an unresolved empirical question in linguistics. But here's the important point in this context: if the distinction between semantic and pragmatic aspects of (3b)'s communicative upshot is controversial among *theorists*, then we could hardly expect it be clear to ordinary speakers. The distinctions between literal and pragmatic ways of communicating information are subtle, esoteric, and often controversial among theorists themselves. In some cases, there are tests that can be used to distinguish pragmatic from semantic content – cancelability, projection, and the like. But such tests can fail to yield decisive verdicts, ordinary speakers are unlikely to be aware of them, and more importantly, they'd have no need for them. As long as we succeed in communicating what we intended, the mode of communication is usually not important except to the theorist with an independent interest in it.

We can sum up our reply to the speaker error worry about semantic vs. pragmatic modes of communication as follows. First, we've observed that the extent to which a metalinguistic analysis attributes error at all depends on what the mechanism for metalinguistic usage turns out to be. So the objection can only have force once currently unsettled empirical questions have been resolved, and only if they are resolved in certain ways. Second, we've suggested that if they *are* a problem for *some* implementations of the notion of metalinguistic usage, then that provides a consideration for our choice among theories of how metalinguistic usage works, and not a consideration against thinking that the usages are metalinguistic in the first place. And finally, we've argued that – even supposing for

the sake of argument that metalinguistic usage involves a distinctive pragmatic mode of communication – the objection about speaker error is not likely to have much force, because speakers are unlikely to have reliable intuitions in this domain, if they have intuitions of the relevant sort at all.

### **3.2. False beliefs about subject matter**

This brings us to the other kind of mistake that a theory of metalinguistic negotiation would potentially ascribe to ordinary speakers. This is the mistake of believing that their disputes do not, in the first instance, concern *representational-level* matters of language and thought, but rather directly address the *object-level* matters that are intuitively at issue. People don't care about the words they use; they care about the things they're talking about. They don't care about 'torture'; they care about torture! (see Cappelen 2018, 174–175).

Our first reply to this worry is to emphasize just how narrow a point it is. Recall that on our view, metalinguistic negotiations are perfectly capable of expressing substantive disagreements well worth arguing about. Moreover, the substantive disagreements they directly express about representational-level issues can be a way of ultimately communicating disagreements about object-level issues. For example: a metalinguistic negotiation concerning the word 'torture' and its application to waterboarding, in the context of a policy debate, can ultimately serve to express disagreements about waterboarding and whether we should do it. A metalinguistic negotiation concerning whether the word 'fish' applies to whales, in the context of a biology classroom, can serve to express disagreements about the location of the biological joints in the universe. And so on. Our account of metalinguistic negotiation is a theory that *vindicates* these ordinary speaker intuitions that they're really disagreeing, and specifically that they're really disagreeing about public policy and waterboarding, about whales and biology, and so on.

What our account doesn't vindicate is the intuition, supposing there is one, that ordinary speakers are expressing these disagreements about object-level issues *directly*, and not via the intermediate step of issues in conceptual ethics that are closely tied to (and perhaps run directly in parallel with) the first-order matters that are intuitively at issue. If this mistaken intuition exists, it concerns how speakers go about expressing their disagreements.

The same considerations raised above concerning beliefs about linguistic mechanism should immediately call into question how plausible it is

that ordinary speakers have the kind of intuitions that we would be forced to describe as mistaken. Is it plausible that ordinary speakers of natural language possess the conceptual toolkit necessary to have beliefs of the kind that would be contradicted by our metalinguistic account, and that speakers would put such tools to use by firmly committing themselves to folk theories about how (object-level only, or object-level *via* representation-level disputes) they express their disagreements? We think it it's not terribly plausible. But grant for the sake of argument that they do. How bad is it that we are forced to ascribe errors of this kind?

Recall the other reason we discussed for why theories that posit speaker error are bad, when they are bad. It is that they insert a wedge between a theory and the data to which that theory is meant to be accountable. A theory of syntax according to which "the cat is on the mat" is ungrammatical, but universally mistaken as grammatical by competent, native speakers of English, has a problem. A theory of phonology on which 'thack' is not a possible word of English but is mistakenly judged to be so by competent, native speakers has a problem. Likewise, a semantic theory of the English word 'disagreement' according to which ordinary speakers misapply it every time they describe ordinary, seemingly functional debates as perfectly good instances of the kind would have a problem.

But of course that's not what our view says, and it's not what our theory is a theory of. Our view provides a way of thinking about how genuine disagreements can be expressed in certain disputes on the assumption that parties to those disputes do not mean the same things by their words. To the extent that ordinary speakers have beliefs about this kind of thing, those beliefs are part of their overall folk-linguistic theory. Those beliefs – not linguistic judgments about how to correctly use an ordinary language term, but folk-linguistic beliefs about the nature and extent of linguistic variation and concept sharing – are no more likely to be accurate, or to be built out of the same concepts that matter to theorists, than folk-scientific beliefs in any other domain.

In fact, our account of metalinguistic negotiation allows the people who do give theories of the meanings of ordinary language terms to stick *closer* to the core linguistic data. After all, the assumption that sameness of meaning is necessary for the expression of genuine disagreement is what leads so many theorists to ascribe meanings to speakers that systematically diverge from those speakers' usage and first-order intuitions. On our account, one can argue that the ancient mariner really does mean by 'fish' exactly what he seems to mean and what a field linguist working with him would theorize that he means: something that includes

both fish and whales. On our account, one can say that the Bush administration official who systematically and sincerely and with complete knowledge of the details of the procedure asserts that “waterboarding isn’t torture” really does mean what she seems to mean by torture: some narrowly circumscribed category including only the most extreme forms of interrogation and punishment.<sup>15</sup>

Attributions of error are not just bad simpliciter. As we’ve noted, in many domains it’s hardly surprising that ordinary speakers get things wrong. Theories attributing error lose credibility when they posit the error they do in such a way as to decrease their accountability to the data that matter. This means that to accuse a metalinguistic analysis of, say, the debate about whether Secretariat is an “athlete”, of attributing too much error gets things precisely backwards. Such an analysis allows for the semantic theory, the theory of the meaning of the word ‘athlete’, to stick *closer* to the data, and in particular to the data suggesting that two speakers with systematically differing dispositions to use the word ‘athlete’ might just mean different things by it. In doing so, it *reduces* the amount of the bad kind of error – the error involved in simply misapplying one’s words. In doing so, it might also ascribe to the speakers a mistaken folk-linguistic belief about the “directness” of language and the particular way in which their disagreement about Secretariat is expressed. But this is an error about exactly the kind of thing about which we expect ordinary speakers to have certain inaccurate views. If a linguistic theory suggests that even with performance errors of the usual kind factored out, speakers radically misapply their own words, that, in general, *is* a problem. If a linguistic theory suggests that speakers have false folk-linguistic beliefs about the nature of language and communication, that, in general, is *not* a problem. To the extent that a theory of metalinguistic negotiation ascribes systematic error to ordinary speakers, it’s error of the latter kind.

Finally, “directness” itself is a tricky notion when it comes to identifying the topic of a disagreement, and in adjudicating between the idea that a dispute is a metalinguistic one vs. a canonical one. Consider the case of torture. When two speakers argue about “whether waterboarding is torture”, many think it is a mistake to claim that the argument is in *any sense* about language. Herman Cappelen, for example, says of such speakers that “their debate, and their disagreement, is independent of how

---

<sup>15</sup>Whether any actual members of the Bush administration were like this is another question. We suspect that however *systematically* they may have asserted such claims, it’s not entirely plausible that those assertions were sincere.

particular words are used. It's about torture, not 'torture'." (Cappelen 2018, 175). But in many cases, the debate that really matters is not about the word 'torture' or about torture. It's about *waterboarding*, and whether we should be doing it. (And how we should treat those who engage in it, etc.) Those are the fundamental normative issues ultimately at stake, at least in many contexts of arguing "about torture". And they will be the kinds of normative issues that many speakers will intuitively take to be at issue here and to really matter.

Arguing about whether waterboarding should be described with the word 'torture' is a worthy activity – provided it takes place against the background assumption that those things that we choose to describe as 'torture' are things that we have strong reason not to do, to discourage others from doing, etc. Without these kinds of assumptions in place, the debate is idle with respect to the political, legal, and ethical issues that are at the heart of the debate. But the exact same point applies if we treat the debate as canonical. Arguing about whether waterboarding *really is torture* is a worthy activity – provided it takes place against the background assumption that those things which *really are torture* are things that we shouldn't do, that we should discourage others from doing, etc. Without that kind of normative assumption in place, the debate is every bit as idle with respect to the practical and ethical issues at the heart of the debate.<sup>16</sup>

To see the import of these kinds of background assumptions, consider the contrast between debates about waterboarding during the Bush administration and during the Trump administration. During much of the Bush administration, it was more or less agreed that the sentence "the United States doesn't torture" should come out true. So a lot rested on how exactly 'torture' was defined, and the Bush Justice Department bent over backwards to find some definition, no matter how strained, that excluded waterboarding. By contrast, Trump has expressed little sincere interest in vindicating the claim that "the U.S. doesn't torture",

---

<sup>16</sup>There could be different debates about what counts as 'torture', – debates concerning sociological or anthropological taxonomy, for example – that hinge on different kinds of issues, and not with normative issues in ethics, politics, and law. But notice that, in these kinds of alternative cases, *other* background assumptions would need to be in place for the debate to not be idle with respect to the issues that participants in *those* debates ultimately care about. Compare: we often argue about what *true democracy* requires, or *morality* requires, in the context of arguing about how we should set up our social-political institutions, or how we should live. We do this against a backdrop of caring about democracy and morality, and thinking they have a kind of normative importance to these latter topics. But a foe of democracy, or an opponent of morality, might still be interested in what democracy or morality requires for *other* reasons – e.g. to show what is problematic about them, or to understand descriptive issues about their particular sociological/historical force. In this regard, consider, for example, (Nietzsche) as well as purely descriptive work about democracies and how they function.

and indeed has argued that without engaging in the most egregious and unlawful abuses of prisoners and enemy combatants, the United States will not be on a “level playing field”. Correspondingly, Trump has shown little interest in defending the practice of waterboarding from being branded as “torture”.

Trump’s willingness to embrace torture brings out the following point. In many arguments about whether waterboarding is torture, it is *common ground* that we shouldn’t do the things we take to be torture (or are willing to describe as ‘torture’), that we should prevent others from doing them, etc. But that isn’t true in all cases. And when it is not true, the debate about *torture* as such will have significantly less importance to the participants. The debate about whether waterboarding is torture has no *intrinsic* interest to people arguing about public policy and the treatment of prisoners, whether it’s analyzed as canonical or metalinguistic.<sup>17</sup> Either way, arguing about whether waterboarding is torture is a way of arguing about *whether we should waterboard*, or about *how we should treat people that do it*, or some other normative issue. That connection – being indirect – only holds when the speakers share the relevant background assumptions. The choice of which of the competing, indirect theories of the debate is better will be a matter of which theory better fits the data of speaker usage and intuition, and which theory fits better into an overall theory of thought and talk.

### **3.3. When do we see the greatest resistance to a metalinguistic analysis?**

Speakers might well resist a metalinguistic analysis of their dispute in some contexts more than others. In previous work, we have suggested that paying attention to aspects of those particular contexts can help explain why that resistance occurs. In other words, it can help explain why speakers make the mistakes they do.

For example, in many contexts, there is an association between (a) the idea that a dispute is about the meaning of words (or other

---

<sup>17</sup>On some specific canonical analyses of the dispute, it could be that the relevant kind of assumptions that we have been discussing are guaranteed to hold. For example, someone might hold that what it is to be “torture” is in part to be the kind of thing to be avoided, or which merits a certain kind of sanction when someone engages in it, etc. Further, they might hold that this follows from the very meaning of the term ‘torture’. If that is right, then an analysis of the dispute as a canonical dispute about torture could yield the result that the speakers directly get at the underlying normative issues that intuitively are at stake in the debate. But even in this case, the “directness” of the dispute is guaranteed by the inclusion of normative / evaluative criteria in the application conditions of the term (or whatever corresponding bits of theoretical apparatus apply), and not simply by the analysis of the dispute *as* canonical.

representational-level matters) and (b) the idea that the disagreement expressed isn't important. Perhaps it is dismissed as a case of 'just arguing semantics'. We think this association is prevalent in much of contemporary American public discourse, for example. It is also true in much of contemporary philosophy. Part of the aim of our work on metalinguistic negotiation is to break down this association. But if this kind of association is in place, it should not be surprising if speakers, in cases where they are convinced that their dispute is about issues that really matter will at least initially resist a metalinguistic characterization of their dispute.

It might also be that speakers resist a metalinguistic characterization of their dispute because they want to resist a conclusion that such a characterization is used to support, or which they suspect it might be used to support.

For example, in previous work we have argued that certain disputes in the legal context are best analyzed as metalinguistic negotiations. In particular, we argued that some of these disputes are metalinguistic negotiations in which the proper use of legal terminology is negotiated as a way of settling what the content of the law in a given jurisdiction *should be*. Yet, in many of these disputes, the speakers are adamant that what they disagree about is *not* what the law should be, but what it is. For example, judges often claim not to be *making* the law but merely interpreting it. This would seem to be an especially tough case for us, as not only are we ascribing error to speakers, but in particular to speakers who pay extremely close attention to language and who explicitly insist on something that looks a bit like the negation of our metalinguistic analysis (For details, see Plunkett and Sundell 2014).

But at the same time that it makes the objection look more pressing, the legal context also provides additional resources for a reply. In many social-historical contexts, there is and has been strong resistance to the idea that judges create law, rather than just apply existing law. Many want to resist it as a descriptive claim about what judges in fact do. Even more want to reject it as a claim about what judges *should* do. That might be a mistaken idea. For example, some legal philosophers have argued that judges will inevitably play a role in the creation of law in virtually all legal systems, given the way in which law is necessarily determined in part by the attitudes and activity of judges themselves, and given that judges are (in virtually all legal systems) legally required to make a judicial decision even in cases where the law runs

out, or when the law is not determinate.<sup>18</sup> But, *if* people have this belief about the proper role of judges, then we should expect them to resist proposals on which they are in fact involved in the creation of law. For this reason, it's entirely understandable that legal actors would resist the idea that the disputes in question are metalinguistic negotiations, given that this interpretation is naturally paired with the idea that engaging in such metalinguistic negotiations is a way of *creating* law.

Other examples come from within philosophy itself. In philosophy, there is often an association between (a) the idea that a dispute is, in the first instance, about the meaning of words (or other representational-level matters) and (b) the idea that we should take a "deflationary" view of the subject matter ultimately targeted by that dispute. This default association might well make sense, given that many philosophers want to combine these two ideas. We have tried to pull them apart. We have argued that metalinguistic negotiation can be a way of ultimately getting at object-level issues that we can take a hardcore realist, non-deflationary view of (see Sundell 2012; Plunkett and Sundell 2013a; Plunkett 2015). But of course not everyone is going to have encountered or been convinced by such arguments. Thus, if all that a philosopher knows is that someone is saying "you are really talking about language in the first instance" and she is skeptical of deflationary views about the relevant subject matter, then it would not be surprising that such a philosopher would (at least initially) resist a metalinguistic characterization of a philosophical dispute she was involved in. She will suspect (perhaps wrongly) that the metalinguistic diagnosis leads to a conclusion that she thinks she has good reason to resist.

How much do these sorts of observations help in mitigating the main worry about speaker error? It will obviously depend more on the details of the specific cases, including the empirical facts about how much speakers resist in one context vs. another. The important point for us is that such further details about a specific context might help bolster our overall response to the worry about speaker error. It also helps underscore a more general point: a proponent of a metalinguistic analysis of a dispute might well have resources to predict ahead of time that speakers will resist such an analysis of that particular dispute (or disputes of a given kind, which that dispute is an instance of).

---

<sup>18</sup>See, for example, (Hart 1961/2012). Hart seems to want to make this as a claim about *all* legal systems, rather than "virtually all" of them, as we have said here. That stronger claim is not crucial in this context. Neither is the claim that this is true of "virtually all" legal systems, as opposed to just *many* ones in the contemporary world.



## 4. Conclusion

We have explored the question of why, exactly, attributions of erroneous beliefs to speakers can be a theoretical cost in developing theories of thought and talk. We've distinguished between speakers misapplying their own words (an attribution which drives a methodologically problematic wedge between a semantic theory and the data to which it is in part accountable), and speakers having false folk-linguistic beliefs about the nature of language and communication (something that should be no more surprising than mistaken folk-theoretic views in other domains). We thus suggested that while error of the first kind is theoretically problematic, error of the second kind is less worrisome. And we observed that the phenomenon of metalinguistic negotiation in fact allows for theorists to posit *less* of the first, problematic, kind, while it attributes error of the second, less worrisome kind only to the extent that we take ordinary speakers to have sophisticated theoretical beliefs of a kind which it's not obvious they have.

## Acknowledgements

An earlier version of this paper was presented at a workshop on Conceptual Engineering hosted by ConceptLab (University of Oslo) in Zermatt, Switzerland in 2017. Thanks to everyone who participated in that workshop for helpful discussion. Special thanks also to Derek Ball, Chris Barker, Ray Briggs, Alexis Burgess, Elisabeth Camp, Herman Cappelen, Matthew Chrisman, Jorah Dannenberg, Andy Egan, Nat Hansen, Stephen Finlay, Justin Khoo, Josh Petersen, Kevin Scharp, Rachel Sterken, Amie Thomasson, and anonymous referees for helpful discussion and comments.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Ball, Derek. Forthcoming. "Relativism, Metasemantics, and the Future." *Inquiry*.
- Barker, Chris. 2002. "The Dynamics of Vagueness." *Linguistics and Philosophy* 25: 1–36.
- Barker, Chris. 2013. "Negotiating Taste." *Inquiry* 56 (2-3): 240–257.
- Björnsson, Gunnar, and Stephen Finlay. 2010. "Metaethical Contextualism Defended." *Ethics* 121 (1): 7–36.
- Burgess, Alexis, and David Plunkett. 2013a. "Conceptual Ethics I." *Philosophy Compass* 8 (12): 1091–1101.
- Burgess, Alexis, and David Plunkett. 2013b. "Conceptual Ethics II." *Philosophy Compass* 8 (12): 1102–1110.

- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- Chalmers, David J. 2011. "Verbal Disputes." *Philosophical Review* 120 (4): 515–566.
- Chierchia, Gennaro. 2004. "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface." In *Structures and Beyond*, edited by A. Belletti, 39–103. Oxford: Oxford University Press.
- Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford: Oxford University Press.
- Finlay, Stephen. 2016. "Disagreement: Lost and Found." In *Oxford Studies in Metaethics Vol. 12*, edited by R. Shafer-Landau, 187–205. Oxford: Oxford University Press.
- Grice, H. Paul. 1989. "Logic and Conversation." In *Studies in the Way of Words*, 1–143. Cambridge: Harvard University Press.
- Hare, R. M. 1952/1963. *The Language of Morals*. Oxford: Oxford University Press.
- Hart, H. L. A. 1961/2012. *The Concept of Law*. 3rd ed. Oxford: Oxford University Press.
- Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34 (1): 31–55.
- Khoo, Justin, and Joshua Knobe. 2016. "Moral Disagreement and Moral Semantics." *Noûs* 50 (2): 109–143.
- Ludlow, Peter. 2008. "Cheap Contextualism." *Philosophical Issues* 18: 104–129.
- Ninan, Dilip. 2010. "Semantics and the Objects of Assertion." *Linguistics and Philosophy* 33 (5): 355–380.
- Plunkett, David. 2015. "Which Concepts Should We Use?: Metalinguistic Negotiations and the Methodology of Philosophy." *Inquiry* 58 (7-8): 828–874.
- Plunkett, David. 2016. "Negotiating the Meaning of 'Law': The Metalinguistic Dimension of the Dispute over Legal Positivism." *Legal Theory* 22 (3-4): 205–275.
- Plunkett, David, and Timothy Sundell. 2013a. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13 (23): 1–37.
- Plunkett, David, and Timothy Sundell. 2013b. "Dworkin's Interpretivism and the Pragmatics of Legal Disputes." *Legal Theory* 19 (3): 242–281.
- Plunkett, David, and Timothy Sundell. 2014. "Antipositivist Arguments from Legal Thought and Talk: The Metalinguistic Response." In *Pragmatism, Law, and Language*, edited by G. Hubb and D. Lind, 56–75. New York: Routledge.
- Rabern, Brian. 2012. "Against the Identification of Assertoric Content with Compositional Value." *Synthese* 189 (1): 75–96.
- Schärp, Kevin. 2013. *Replacing Truth*. Oxford: Oxford University Press.
- Schroeter, Laura, and Francois Schroeter. 2014. "Normative Concepts: A Connectedness Model." *Philosophers' Imprint* 14 (25): 1–26.
- Silk, Alex. 2016. *Discourse Contextualism: A Framework for Contextualist Semantics and Pragmatics*. Oxford: Oxford University Press.
- Sundell, Tim. 2011. "Disagreements about Taste." *Philosophical Studies* 155 (2): 267–288.
- Sundell, Tim. 2012. "Disagreement, Error, and an Alternative to Reference Magnetism." *Australasian Journal of Philosophy* 90 (4): 743–759.
- Sundell, Tim. 2016a. "Eligibility and Ideology in the Vat." In *The Brain in a Vat*, edited by S. Goldberg, 226–250. Cambridge: Cambridge University Press.
- Sundell, Tim. 2016b. "The Tasty, the Bold, and the Beautiful." *Inquiry* 59 (6): 793–818.
- Thomasson, Amie L. 2016. "Metaphysical Disputes and Metalinguistic Negotiation." *Analytic Philosophy* 57 (4): 1–28.

- Thomasson, Amie L. Forthcoming. "A Pragmatic Method for Conceptual Ethics." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett.
- Yalcin, Seth. 2014. "Semantics and Metasemantics in the Context of Generative Grammar." In *Metasemantics*, edited by A. Burgess and B. Sherman, 17–54. Oxford: Oxford University Press.