Topoi
https://doi.org/10.1007/s11245-023-09941-2

Varieties of Metalinguistic Negotiation

David Plunkett[1] **· Timothy Sundell**[2]

Accepted: 1 June 2023

Abstract
In both co-authored and solo-authored work over the past decade, we have developed the idea of "metalinguistic negotiation". On our view, metalinguistic negotiation is a type of dispute in which speakers appear to use (rather than explicitly mention) a term in conflicting ways to put forward views about how that very term should be used. In this paper, we explore four possible dimensions of variation among metalinguistic negotiations, and the interactions among those dimensions. These types of variation matter for understanding the nature, and the potential range, of the phenomenon of metalinguistic negotiation. As an illustration of the latter, we argue in our concluding section that understanding the full range of forms that metalinguistic negotiations can take has implications for debates about the "implementation" of conceptual engineering proposals.

Keywords Metalinguistic negotiation · Metalinguistic disputes · Metalinguistic usage · Conceptual ethics · Conceptual engineering

1 Introduction

In both co-authored and solo-authored work over roughly the past decade, we have developed the idea of "metalinguistic negotiation".[1] On our view, metalinguistic negotiations are a type of dispute in which speakers appear to use (rather than explicitly mention) a term in conflicting ways to put forward views about how that very term should be used. We have argued that a number of disputes of interest to philosophers, as well as some disputes amongst philosophers themselves, might best be thought of as metalinguistic negotiations. We have advanced this view with respect to disputes about morality, law, aesthetics, and metaphysics, among other topics. In our work thus far, our aim has rarely been to establish that some particular dispute definitely is a metalinguistic negotiation. Rather, we have generally aimed to show that a metalinguistic analysis of a range of disputes in a particular area of discourse is plausible, and to explore the dialectical significance of that for work in the relevant area of philosophy.

To illustrate, we have argued that prominent "disagreement-based" arguments in philosophy lose much of their force once the idea of metalinguistic negotiation is firmly in view.[2] In general, these arguments begin with an observation that some range of disputes seems to express genuine disagreements, and they end with the conclusion that parties to those disputes mean the same things by certain key words. But that conclusion is supported only to the extent that we can rule out analyses of a certain sort for the relevant range of cases: namely, analyses on which genuine disagreements are expressed despite variation in the meanings of the relevant words. The truth of this type of analysis does not have to be conclusively demonstrated in order to establish the point that such disagreement-based arguments cannot move, without substantial further argument, from a premise about genuine disagreement to a conclusion about sameness of

✉ David Plunkett
david.plunkett@dartmouth.edu

Timothy Sundell
timothy.sundell@uky.edu

[1] Philosophy, Dartmouth College, Hanover, NH, USA

[2] Philosophy, University of Kentucky, Lexington, KY, USA

[1] For the relevant co-authored work summarized in this paragraph, see Plunkett and Sundell (2013a, b, 2014, 2021a, b). We both develop and expand on the idea of metalinguistic negotiation in solo-authored work, including in Plunkett (2015, 2016), Sundell (2011a, b, 2016, 2017). We also further develop our views in recent co-authored work with Rachel Sterken (2023).

[2] See Plunkett and Sundell (2013a, b).

Published online: 05 August 2023

meaning. Rather, to establish this point, one needs to show merely that an analysis involving semantic variation is plausible for the relevant cases. As we have understood them, metalinguistic negotiations are an example of this latter type of analysis. Given the importance of disagreement-based arguments across a number of philosophical subfields—Terence Horgan and Mark Timmons's "Moral Twin Earth" argument in metaethics, and Ronald Dworkin's argument about "theoretical disagreement" in philosophy of law, to name just two—this result can matter significantly for the subfields in question.[3]

This dialectical aim—calling into question the move from genuine disagreement to sameness of meaning—has impacted our presentation of metalinguistic negotiation in a number of ways. In particular, we have advanced a highly schematic account of what metalinguistic negotiations are and how they operate. For example, we have generally presented metalinguistic negotiation as a *pragmatic* phenomenon, where the "metalinguistic" usage of a term conveys information that isn't part of the semantic content that is asserted. However, we have emphasized that other accounts—dynamic semantic accounts, for example—are possible, where the conveyed information is part of the semantics, or where a sharp semantics/pragmatics divide isn't assumed.[4] We have remained neutral on this and other issues for the following reason: for most of our argumentative purposes, it simply doesn't matter how exactly the phenomenon of metalinguistic negotiation is cashed out. When it comes to isolating the problematic assumptions at work in disagreement-based arguments for sameness of meaning, for example, what matters is that the phenomenon of metalinguistic negotiation is real, and that the disputes in question are plausible candidates for such an analysis. Firm commitments with respect to questions about mechanisms aren't necessary in such an argumentative context, and would in fact distract from the philosophical goals of these discussions.[5]

Because of this dialectical context, our discussions about "metalinguistic negotiation" have tended to focus first on a relatively small set of examples that we use for illustration, and then on whatever kind of dispute is most relevant to the philosophical domain we address in a given context. For purposes of illustration, we have repeatedly used examples about whether horses can be "athletes", whether waterboarding is "torture", whether chili is "spicy", or whether some office is "cold", moving on from those illustrations to disputes relevant to a given philosophical subfield—whether Vegemite is "tasty", whether a given action is "moral", whether whales are "fish", whether something is "the law" in a given jurisdiction, etc. This general strategy makes sense when it comes to presenting metalinguistic negotiation in the dialectical contexts of previous papers. But, collectively, this kind of presentation has the potential to give an impoverished picture of the full range of metalinguistic negotiations, or the misleading impression that the account we give cannot accommodate cases that haven't made their way into our previous work.

In this paper, we aim to step back from the dialectical aims that have shaped these previous discussions, and explore more broadly the range of cases that are plausibly treated as instances of metalinguistic negotiation. Specifically, we'll address four ways that metalinguistic negotiations can vary from each other, and the interactions among and implications of those types of variation. This variation matters for understanding the nature, and the potential range, of the phenomenon of metalinguistic negotiation. The four types of variation we discuss are:

(1) Variation in speakers' *motivations* for entering into a metalinguistic negotiation.
(2) Variation in the *audience(s)* that the participants in the metalinguistic negotiation are trying to communicate with.
(3) Variation in the *scope* of the normative claims the speakers advance as part of the metalinguistic negotiation.
(4) Variation in the *kind* of normative or evaluative claims the speakers advance in the metalinguistic negotiation.

In what follows, we start by providing an overview of our account of metalinguistic negotiation in general. We then turn to each of these dimensions of possible variation in turn. Each kind of variation is one that we find in many other kinds of disputes as well. We thus aren't making any claim that the presence of such variation is unique to metalinguistic negotiations. We discuss these dimensions of variation in order to advance our understanding of metalinguistic negotiations, and the roles that they might play in discourse across a range of domains.

We have a few different overarching goals in exploring these types of variations, and their interactions.

Our first, and primary, goal is to better understand the phenomenon of metalinguistic negotiation, including both the ways metalinguistic negotiations can differ from each other and what metalinguistic negotiation is more generally. The possible differences among metalinguistic negotiations that we discuss are not entirely novel to our exploration in

[3] For further discussion about the metaethics case, concerning the argument put forward in Horgan and Timmons (1993), see Plunkett and Sundell (2013a), and for further discussion focused on philosophy of law, regarding the argument put forward in Dworkin (1986), see Plunkett and Sundell (2013a).

[4] See Plunkett and Sundell (2021b).

[5] For further discussion, see Plunkett and Sundell (2021b).

this paper. Indeed, much of what we argue and observe here simply makes more explicit ideas that run throughout our previous work. But, in making these issues explicit—and in organizing them in a new way, disengaged from particular arguments rooted in other subareas of philosophy—we hope to deepen general appreciation and understanding of metalinguistic negotiation. As we discuss, thinking through the differences between types of metalinguistic negotiation allows us to highlight kinds of metalinguistic negotiation that aren't foregrounded in our previous discussions. It also allows us to raise important questions about how unified a linguistic category "metalinguistic negotiation" is, and, in particular, how unified the underlying linguistic mechanisms involved in the relevant disputes are.

Second, we hope to show that appreciating the full range of metalinguistic negotiations can help us better understand the capacious nature of the phenomenon of metalinguistic negotiation on our account, and what that account is (and is not) committed to. Doing so can help blunt certain critiques of our account that one might otherwise be drawn to.

Finally, we think there is some possibility that better understanding the full range of metalinguistic negotiations, or certain kinds of metalinguistic negotiations that have not been the focus of previous work, may prove relevant to debates within particular subfields of philosophy. Discussions of disagreement-based arguments for sameness of meaning are not the only place where issues about the nature of disagreement, or how we give voice to our disagreements in conversations, matter. With this in mind, we think it could easily turn out that metalinguistic negotiations of kinds that we haven't previously focused on, but which emerge from our discussion here, might prove philosophically illuminating in different philosophical contexts. Our aim in this paper is not to explore such potential upshots in detail. We do, however, offer preliminary thoughts at the end of the paper on one place our discussion may have an impact: namely, the issue of linguistic change, and how this connects to questions about "conceptual engineering".

2 Metalinguistic Negotiation: An Overview

In order to set the stage for the discussion below, we first need an overview of what "metalinguistic negotiation" is, as we understand it. In this section, we provide that overview. We don't cover every detail of the analysis developed in previous work, but rather focus on those aspects of the view that matter most for the discussion in this paper.[6] Having said that, readers already familiar with the framework may prefer to skip this section.

On the account that we develop, metalinguistic negotiations are a certain kind of dispute. By 'dispute', we mean a linguistic exchange (such as a conversation) that appears (e.g., to a theorist considering the exchange) to evince a disagreement. By 'disagreement', we mean a kind of rational conflict in mental states, such as one person believing a proposition that another denies, or (perhaps) one person having a plan that conflicts with the plans of another person in the right way. These definitions are stipulative, and are meant to resonate (if only imperfectly) with how philosophers have often used these terms in recent work, while helping to zero in on the relevant linguistic and philosophical considerations.

With these understandings of 'dispute' and 'disagreement' in hand, we can raise the following question about a given dispute: does it really express a disagreement or not? If the answer is "yes", we can go on to ask what that disagreement is about, and by what linguistic mechanisms the disagreement is expressed. Note that it could be that multiple disagreements are expressed in a single exchange. We think this is often the case. If so, these further questions pertain not to a single disagreement, but to the range of disagreements expressed. We can also go on to raise questions about how these disagreements are related to each other, including which ones depend on which others, which are most salient to the speakers, and which are expressed explicitly or implicitly in the utterances making up the exchange.

For disputes that do in fact express disagreements, we can ask, "Is this disagreement expressed via the literal semantic content of the words the speakers use, or through other mechanisms?" We ask this bearing in mind, as emphasized above, that the disagreement immediately expressed in the exchange—whether it's expressed semantically or pragmatically—may not be the only thing the speakers disagree about, or even the most important thing. In much of the literature we were responding to in earlier work, it was assumed that in a typical case the most important conflicting information in a dispute expressing a genuine disagreement will be the semantic content of the expressions the speakers use, as opposed to information that is conveyed via implicature, presupposition, or other pragmatic mechanisms. With a nod to that assumption, we call disputes where the conflicting information is expressed semantically *canonical* disputes, and disputes where the disagreement is expressed via pragmatic mechanisms *non-canonical* disputes.

Within the category of non-canonical disputes, we introduce what we call *metalinguistic disputes*. Metalinguistic disputes are disputes where speakers appear to use (rather than mention) a term to put forward views about that very term. This is a kind of usage that, following Chris Barker,

[6] Our discussion below draws together ideas presented in the works cited in the first footnote, while closely following the basic line of presentation from Plunkett and Sundell (2013a).

we call a *metalinguistic usage*.[7] Within the category of metalinguistic disputes, we make the following distinction. In some metalinguistic disputes, the speakers disagree about *descriptive* questions concerning the term at issue—how people in fact use the term, what the term in fact means, etc. We call these *descriptive* metalinguistic disputes. In other metalinguistic disputes, speakers focus on *normative* questions about the term at issue—how people *should* use that term, what the term *should* mean, etc. We call these *normative* metalinguistic disputes. We introduce the phrase *metalinguistic negotiation* as a synonym for 'normative metalinguistic dispute'.

Metalinguistic negotiations are therefore characterized by two distinctive factors: (1) the mechanism by which a disagreement is expressed (namely, the mechanism Barker calls a "metalinguistic usage" of a term) and (2) the kind of issue those disagreements are, in the first instance, concerned with (namely, normative issues about how some linguistic expression should be used).

In some metalinguistic negotiations, speakers agree on the standing meaning of a given term, but disagree about how it should be used in a given context. For example, speakers might agree on the context-invariant meaning of a relative gradable adjective (like 'tall', 'rich', or 'cold'), but disagree about what the context-specific threshold should be (e.g., for height, financial wealth, or temperature). Similarly, they might disagree on how to make a vague term more precise. In other metalinguistic negotiations, the conventional meaning itself—what we might think of as the term's *character*—might be at issue. To illustrate, two speakers might disagree about what 'hotel', 'shoe', 'violence', or 'sandwich' should mean for a given context. Terms like these are not usually taken to be context-sensitive in the sense that 'tall' or other relative gradable adjectives are. So when speakers advocate for what such terms should mean in metalinguistic negotiations (as we think they can), it won't be plausible to analyze them as having different views about how to set a context-sensitive parameter in the same way they might for a term such as 'tall'. Instead, an account where they are targeting the term's character will often be more plausible. Indeed, even if terms like these turned out to be context-sensitive in a more traditional sense (along the lines of 'tall'), speakers might still have divergent views about what the context-invariant meaning of those terms should be. After all, it's possible to imagine speakers having divergent views about the context-invariant aspects of the meaning of 'tall', 'rich', or 'cold'. (They could advocate for a meaning on which those terms reflect positions on scales *other* than height, financial wealth, or temperature—a scale of "richness" that reflects purely emotional fulfillment, for example.)

The normative issues about language and concepts at issue in metalinguistic negotiations can be thought of in terms of what Alexis Burgess and one of us (Plunkett) have dubbed *conceptual ethics*.[8] Drawing on that work, we use this term to refer to a cluster of normative and evaluative issues about thought and talk. If we assume that words have meanings, and that there are such things as concepts, we can gloss issues in "conceptual ethics" as having to do with what we should mean by our words, and why, and which concepts we should use, and why. Related evaluative issues include questions about which meanings of words are better or worse, and what would improve our concepts.

The term 'conceptual ethics' is meant as a helpful shorthand for these normative and evaluative issues about thought and talk—issues that come up throughout different subareas of philosophy, and which have done so for a long time. The term 'conceptual ethics' should not be read as taking a substantive stand on what these issues ultimately involve. In particular, the term isn't meant to suggest that the key issues here are about *concepts* as opposed to other items in thought and talk (e.g., conceptions, words, etc.). Which items a given theorist thinks will be on the list here, and how they relate to each other, depends in large part on which theories of thought and talk that theorist is working with, and why. Neither is the term meant to suggest that the key norms or values involved in conceptual ethics are those that are fundamentally "practical", such as those that find their primary home in moral and political philosophy. Indeed, many working in what we take to be "conceptual ethics" emphasize norms or values that (at least on some ways of distinguishing them) are largely "epistemological" (such as helping to foster understanding) or "metaphysical" (such as "carving reality at its joints").[9] The fact that different philosophers writing in or about "conceptual ethics" appeal to different norms or values reflects the following, broader fact: philosophers disagree about many issues related to conceptual ethics, including how, when, and why a given concept should be used in a given context, or what makes one concept better or worse than another.[10]

Disagreement about general issues in conceptual ethics is not peculiar to philosophers. Ordinary speakers will often have differing views on general issues in conceptual ethics (such as what norms and values matter for settling word

[7] Barker (2002).

[8] Burgess and Plunkett (2013a, b).

[9] On the understanding of "conceptual ethics" we are working with, conceptual ethics is sometimes tied to "conceptual engineering", which (put roughly) connects conceptual ethics to work on the design and implementation of proposed conceptual changes. For further discussion of this connection, see Burgess and Plunkett (2020) and Cappelen and Plunkett (2020). Later in this paper, we turn to some issues about the implementation aspect of conceptual engineering projects.

[10] For connected discussion, see Burgess and Plunkett (2013a, b), and Cappelen and Plunkett (2020).

choice), even if those views are often inchoate or tacitly held. As we have discussed in previous work, these differences in views are likely to be reflected in metalinguistic negotiations. Our general account of metalinguistic negotiation is a descriptive one, which we think philosophers with a wide range of normative views in conceptual ethics can accept. Our account thus doesn't take a stand on which foundational views of conceptual ethics are correct. Given the kind of descriptive account it is, the key thing is that it is able to capture the kind of variation in views that speakers bring to the table, as well as the dynamics of how those views play out in given conversations. In other work, we have made the case that it can.[11]

In addition to making room for a variety of views in conceptual ethics, our view also allows for the possibility that the standards an individual speaker thinks matter can vary significantly from context to context, even when the same term's meaning is being negotiated. As we have emphasized in other work, we think there is good reason to expect this kind of variation.[12] If a speaker is sympathetic to the notion of objective "joints" in nature, they might evaluate a particular metalinguistic negotiation in a scientific context with respect to how well the proposed categorizations match those joints. In a non-scientific, ordinary practice context, the very same term might play a different role, and proposed categorizations might be evaluated by that speaker with respect to an entirely separate, practical, set of standards. In some contexts, a speaker might take one concept to be better than another in virtue of reflecting our moral, ethical, or political values. In the case of metalinguistic negotiations engaged in to pass the time, with no particular stakes in mind, a speaker might think that the proposals might fail to be evaluable as better or worse at all.

Our view is consistent with speakers being *wrong* about the standards by which they judge proposals about what to mean by their words, or by which they evaluate the goodness of a concept. For example, a speaker might believe that a certain form of consequentialism is appropriate for assessing all issues in conceptual ethics, and it could turn out that consequentialism in conceptual ethics is wrong. Or a speaker might have badly mistaken views about what hangs on a particular word choice, such that she overlooks important moral or epistemic considerations relevant to that choice. At the same time, we think it is plausible that the correct account of conceptual ethics will involve an important kind of contextual variation, reflective of variation in speakers' beliefs and aims, such that, for example, it's better to use the term 'fish' one way in a scientific context and another way when creating a restaurant menu. While we don't take a stand on this issue, we think it's plausible that how speakers should use a given term in a given context is partly determined by facts about speakers' *aims* in the relevant context.[13]

As this discussion makes clear, we have presented metalinguistic negotiations in schematic terms. While our account aims to identify what metalinguistic negotiations have in common, and how they differ from other kinds of disputes, we have consciously aimed to emphasize that "metalinguistic negotiation" is a broad category, involving a number of significant further cuts that can be worth making. Moreover, although we have presented the notion of a "metalinguistic usage" as a pragmatic mechanism, nothing broader hinges on that assumption, nor indeed on any particular assumptions about the underlying linguistic mechanism. It could turn out that important elements of metalinguistic usage are semantic, or that the phenomenon Barker calls a "metalinguistic usage" of terms is best understood within the framework of dynamic semantics, where the divide between semantics and pragmatics plays out differently than in non-dynamic frameworks.[14] We think of these as important questions in linguistics, philosophy of language, cognitive science, and connected areas.[15] But, for the core of our own work thus far, what's mattered is getting the basic phenomenon of "metalinguistic negotiation" squarely in view, and understanding that it shows up in parts of ordinary life as well as in key parts of theoretical discourse, in philosophy and beyond.

3 Motivations

Now that we have an overview of metalinguistic negotiation, we can turn to four further factors by which metalinguistic negotiations can vary, the dimensions of variation we introduced at the start of this paper.

Let's start with the issue of the motivations (either explicit or not) that speakers have for engaging in metalinguistic negotiation.

Why would speakers be motivated to engage in metalinguistic negotiations? We think that there is no *single* answer

[11] See, for example, our work in Plunkett and Sundell (2013a, 2021a, b).

[12] For more discussion of the points in what follows, see Plunkett and Sundell (2013a, b), Sundell (2011a, 2016), and Plunkett (2015).

[13] For connected discussion of this kind of idea, see Burgess and Plunkett (2013b), Plunkett and Sundell (2013a), Sundell (2011b), Plunkett (2015), Haslanger (2000), and Thomasson (2020).

[14] For example, consider the framework that Barker develops in Barker (2013). For further discussion, see Plunkett and Sundell (2021b).

[15] For some different recent takes on this issue (or closely connected ones, about similar kinds of disputes), see Mankowitz (2021), Thomasson (2016), Khoo (2020), Belleri (2017), Kocurek et al. (2020), Bolinger (2022), and Soria-Ruiz (2023). We briefly discuss some of our own ideas on this issue in Plunkett and Sundell (2021b).

to this question, and that none should be expected. For any kind of dispute—canonical or non-canonical, in conceptual ethics or in any other domain—speakers might have a range of motivating reasons for engaging in that dispute. In short, there's no reason to expect that people across different contexts, with different psychologies, with different conversational partners, etc. will have uniform motivations for engaging in metalinguistic negotiations.[16]

To illustrate, consider Peter Ludlow's 'athlete' case, which we have repeatedly discussed in our earlier work.[17] In this case, two speakers argue on sports radio about whether Secretariat the horse belongs on a list of the greatest athletes of the twentieth century. As the dispute unfolds, it becomes clear that one speaker doesn't think horses can ever be "athletes", let alone great ones, while the other thinks they can be. Drawing on Ludlow, we suggest that the two speakers are plausibly read as offering not different substantive views about what falls under a shared common concept they both deploy (call it ATHLETE), but rather views about which of a range of related "athlete"-type concepts is appropriate, and thus what the character of the term 'athlete' should be, for the relevant context.[18] In other words, we suggest the dispute is a metalinguistic negotiation with speakers expressing different normative views about what the term 'athlete' should mean, or at least how the term should be used for purposes of their conversation.

Consider the motivations of the speakers in this case. Why should they care, even tacitly, about this issue in conceptual ethics? For any of a wide range of reasons. Whether horses count as "athletes" can tie in to a range of ethical and political issues, such as the treatment of non-human animals, or how (and why) we recognize certain forms of excellence among humans. Speakers might care about those issues independently, and see the relevant issues in conceptual ethics about 'athlete' and the range of "athlete"-type concepts at play as tied in to those issues. Alternatively, the speakers might feel that certain practical matters hang on the dispute. For example, one or both of the speakers might hope that some listeners become convinced of their views, which might in turn contribute to some change in practices around sports, horse racing, the treatment of non-human animals, etc. The speakers could be wrong of course. People can overestimate their ability to make a practical impact. But it's also easy enough to imagine—especially if the radio show has a large audience—that the parties to this dispute are justified in believing their statements could have a practical impact of some kind. They might (perhaps rightly) think that the practical impact of the dispute will in part be tied to which of the two conversational adversaries makes the more convincing case, or to which case resonates most with the listeners to the radio show, etc.

Now contrast the 'athlete' case with another that has come up in our previous discussions—a dispute about the spiciness of chili.[19] Two people are discussing how to write up the menu for a restaurant they are opening together. It's just the two of them, talking privately about what counts as "spicy", after tasting a sample of the chili they intend to serve at their restaurant. One speaker describes it as "spicy", while the other insists that it is "not spicy". On our account, cases like this are plausibly analyzed as ones where speakers agree on what the term 'spicy' means in general, but disagree about how a context-sensitive parameter for this term should be set in their context.

In contrast to the Secretariat case, these speakers can be seen as attempting to coordinate on a joint activity, and might well be doing so without taking a fundamentally adversarial, but rather a cooperative, stance towards one another. How a restaurant labels things on a menu can matter for what kind of clientele it attracts, which items on the menu people order, etc. If we assume that the people involved in the dispute have a shared interest in running a successful business together, they might well want to coordinate on the relevant threshold of 'spicy'. And they might go in to the dispute with every expectation that they will in fact manage to coordinate after enough discussion.

Contrast that last point with the Secretariat case. It's easy in that case to imagine each speaker believing that the other is likely to keep digging in, and is unlikely ever to change her views on the topic, or at least not on the basis of this particular dispute. That's a reflection of the fact that, in the Secretariat case, coordination between the speakers on the meaning of a term—let alone coordination on a more general joint project—simply isn't a core part of what they are trying to achieve.

We've seen so far that speakers can enter into metalinguistic negotiations to express theoretical views about some topic, or to attempt to effect some practical change. They can go into these negotiations in an adversarial or cooperative spirit. And they can go in with the expectation that coordination will actually be achieved, or the expectation that coordination is unlikely, with some alternative set of goals—perhaps pertaining to listeners outside the bounds of

[16] For previous work where we emphasize this point, see Plunkett and Sundell (2013a, b, 2021a).

[17] For Ludlow's original discussion of the case, see Ludlow (2008). For some of our previous discussion of this case, see Plunkett and Sundell (2013a, b).

[18] In this paper, we use small caps (e.g., CAT) to name concepts, italics (e.g., *cat*) to introduce terminology or for rhetorical stress, single quotes (e.g., 'cat') to mention linguistic expressions, and double quotes (e.g., "cat") for quoting other authors, "scare quoting", simultaneous use and mention, and other informal uses.

[19] See Plunkett and Sundell (2013a).

their immediate conversational context—serving to justify the effort. (This last point ties into the question, discussed below, of which audiences are involved in a given metalinguistic negotiation.) These observations arise from just two examples, and indeed two examples that we have already discussed at length in previous work. But they nonetheless serve to bring out just how much speaker motivations can differ from one metalinguistic negotiation to another.

Once we start appreciating this basic point, it's easy to explore further motivations that speakers might have for engaging in metalinguistic negotiations. Here are just a few:

- A speaker might aim to coordinate with her interlocutor on uses of terminology to aid in joint projects of inquiry, in order to allow that inquiry to proceed more smoothly going forward.
- A speaker might aim to formulate theories that "carve reality at its joints", and, in turn, advocate for concepts and words she believes to have that feature.[20]
- A speaker might be trying to win a political campaign, or achieve any number of other practical ends other than achieving successful inquiry.[21]
- A speaker might want others to understand her and what she is trying to communicate with a given term, even if she doesn't particularly care about whether she succeeds in coordinating on that exact meaning or usage with her interlocutor, or with anybody else.
- A speaker might want to put forward views in conceptual ethics simply to "bear witness" to the truth as she sees it, even if her claim, or her advocacy for some view in conceptual ethics, is unlikely to have any practical effect on anyone in the conversation or out of it.
- A speaker might enter into a metalinguistic negotiation in an attempt to browbeat, intimidate, or silence her interlocutor, and aim to do so by metalinguistically advocating for usages that are at cross purposes to that person's goals, or that undermine that person's standing in some way.[22]
- A speaker might metalinguistically advocate for some usage as a way of demonstrating a certain kind of social identity or group membership by using (and promoting the usage of) the right terminology in the right way.[23]
- A speaker might be trying to advance certain normative views, while dodging accountability for doing so by avoiding straightforward assertion of those views.
- A speaker might enter into a metalinguistic negotiation *just for fun*, debating for example whether a hot dog is a "sandwich", or whether water—the water itself, not the objects it touches—is "wet", all with no particular expectation of coordination, and no particular stakes in mind.[24]

This is just a partial list of candidate motivations we think parties to a metalinguistic negotiation might have. The full list here would be massive, and open-ended.

It should be emphasized that, as in many domains of behavior and psychology, the motivations that in fact psychologically cause speakers to engage in disputes might not be transparent to them. Nor would speakers necessarily agree with an accurate description of their motivations if they were presented with such psychological descriptions. This is a reflection of the more general fact that people can be mistaken about their own mental states, and in their assessment of theories of those mental states. In some metalinguistic negotiations, then, we should expect speakers to be well aware of their motivations, whereas in other cases we should not.[25]

Four further points about speaker motivations in a metalinguistic negotiation are worth emphasizing.

First, recall the earlier idea that a speaker's aims in a given context play a role in determining the normative facts in conceptual ethics about what she should mean by her words. If that's right, then the issue of which motivations a speaker has for engaging in some metalinguistic negotiation is closely tied to the issue of the standards by which we should assess a speaker's metalinguistic proposals, and the normative views in conceptual ethics she puts forward in that dispute.

[20] Versions of this goal, and the previous one, are arguably at the core of a number of cases we have discussed at length in previous work; including metalinguistic disputes about how best to use philosophical or scientific terms. See Plunkett and Sundell (2013a), Plunkett (2015), and Sundell (2011a).

[21] In much of our previous work we have emphasized the importance of differences between these kinds of aims and the aims of engaging in successful inquiry. See, for example, our discussion in Plunkett and Sundell (2013a).

[22] Consider Crocodile Dundee's statement, insulting a blade-wielding interlocutor who had attempted to appear threatening: "That's not a knife. *This* is a knife." For recent work that emphasizes the kinds of practical issues about domination and power in metalinguistic negotiation (or connected disputes involving conceptual ethics and conceptual engineering), see Shields (2021) and Podosky (2022). These issues are also prominent in Ludlow's discussion in (Ludlow 2014).

[23] For connected discussion, see Davies (2021). See also related ideas in Eidelson (2023), drawing on Nunberg (2018).

[24] We discuss this more in Plunkett and Sundell (2021a).

[25] This is a point that we have explored in greater detail in previous work. See Plunkett and Sundell (2014, 2021b). We emphasize in that earlier work that it's consistent with our account not only that speakers engaged in a given metalinguistic negotiation not be aware of it, but indeed that they might vigorously resist such a characterization. As we have argued, there are theoretical resources available not only to accommodate speakers' lack of awareness, but also to provide a deflationary account of that resistance. See Plunkett and Sundell (2021b). For recent critical discussion of our views here (which it is beyond the scope of this paper to address), see Abreu (2023) and Odrowąż-Sypniewska (2023).

Second, we should distinguish the motivations for *engaging* in a given dispute from the motivations we might have for *forming views on a topic* that the dispute is about. One might have all sorts of views in conceptual ethics about a given term that one simply isn't motivated to express in a given conversation—just as one might have views on any other topic (whether biology, political philosophy, or the behavior of one's friends) that one isn't motivated to express in a particular conversational context. In general, the motivations here can come apart from a speaker's overall beliefs or attitudes in significant ways. For example, as noted above, it might be that one engages in a metalinguistic negotiation in order to browbeat, intimidate, or silence someone. Metalinguistic negotiation might be just one strategic means to an end here, such that the motivation for engaging in it may have little or nothing to do with a given speaker's actual views in conceptual ethics.

Third, note that this point about motivation ties into a connected normative one: namely, that the justification for engaging in a given dispute should be distinguished from the justification for having a certain belief (or other attitude), or engaging in reflection or inquiry on that topic.

Finally, we conclude this section with an observation about how these points relate to the choice of 'metalinguistic negotiation' as a label for the phenomenon we've aimed to characterize.[26] As we emphasized, the term 'metalinguistic negotiation' is introduced as a stipulative, (hopefully) helpful shorthand for referring to those disputes that we also characterize as "normative metalinguistic disputes". Those disputes are distinguished by two key things: that they involve competing metalinguistic usages of some expression, and that they focus on certain normative issues about words and concepts (issues in "conceptual ethics"). Our view therefore doesn't require that *metalinguistic* negotiations play out in a way that mirrors the nature of negotiations more generally. As the term is stipulative, an understanding on which ordinary "negotiations" play out in some way that fails to parallel what we see with normative metalinguistic disputes poses no particular threat to the idea of "metalinguistic negotiation".

In fact, however, there is good reason to think that negotiations in the broadest sense *do* play out in ways that mirror what we see with normative metalinguistic disputes. Negotiations in general can be entered into in good faith or in bad faith. They can involve participants who are fairly matched in power or social standing, or they can involve radically unfair or unjust power imbalances. They can be entered into in the spirit of cooperation, or in a thoroughly adversarial mindset. And they can be carried on with every expectation of reaching a reasonable conclusion, or they can be carried on with no particular expectation of reaching a result that the parties involved will agree on. (For example, when carried on for show, for ritual, or out of blind optimism.) We think there are, in fact, lots of ordinary negotiations that aren't aimed at coordination, where cooperation really isn't feasibly on the table, and is not even a kind of regulative ideal. Just like with metalinguistic negotiations, we enter into any other negotiation for a wide range of reasons, with a wide range of goals and expectations.

At the same time, we don't rely on these parallels to motivate our account, and there could be disanalogies as well. It could turn out that negotiations not aimed at achieving cooperation, coordination, or agreement are best thought of as in some sense "marginal" relative to those that are—or that they are explanatorily parasitic on those that are. (Of course that could be advanced as a view about metalinguistic negotiations as well.) If such a view proved plausible about negotiations in general but implausible about metalinguistic negotiations, or vice versa, then there would be an important difference between those things that are properly called "negotiations" and what we have labelled "metalinguistic negotiations". The key point for us is that—however these debates about "negotiation" in general play out—aligning with ordinary notions or philosophical analyses of "negotiation" more generally is not a theoretical commitment of our account of normative metalinguistic disputes.

4 Audience(s)

We now turn to the second of the four types of variation: variation in the intended *audience(s)* for a given metalinguistic negotiation.

Just as speakers can have a range of conscious or unconscious motivations for engaging in a metalinguistic negotiation, so too might they have a range of perceived or intended audiences for a given metalinguistic negotiation.[27] There might be a difference in intended audience across parties to a single metalinguistic negotiations. Even a single speaker might have multiple audiences in mind, perhaps in ways that won't be obvious to others involved in or overhearing the dispute.[28] To see some of the wide range of options here, consider again the Secretariat and 'spicy' examples from the last section, along with a new, third case.

[26] For connected discussion, see Plunkett and Sundell (2021a).

[27] For previous discussion, see Plunkett and Sundell (2013a) and Plunkett (2015).

[28] This point connects to the phenomenon of "code words" (sometimes also discussed as "dog whistles"), and suggests ways in which metalinguistic negotiation might connect to some instances of the use of such words. For some of the recent discussion here on "code words", and how their use ties into propaganda and other important socially important types of speech, see Khoo (2017), Saul (2018), Stanley (2015), and Quaranto (2022).

In the Secretariat case, the dispute is on sports radio. So the audience isn't just the speakers engaged in the dispute: it's also everyone who is listening to the sports radio show. Of course the speakers engaged in the dispute will not know who those specific individuals are. But they will have some sense that there are people listening to the show, and likely some sense of the basic demographics of that audience. Each speaker may in fact be much more concerned with the effect of her utterances on that audience, as compared to her actual interlocutor—whom she may have no particular expectation of convincing or of coordinating with.

Contrast that with the 'spicy' case. While the two restaurant entrepreneurs might well be concerned about the audience that will eventually read their menus, the immediate audience is just them: the two people involved in the dispute. As we have constructed the case above, no one else is around the restaurant to hear their dispute. If these speakers hope to effect some change with their utterances, the immediate change can only be in the very person they're arguing with. At the same time, given the way we set up the context, if the two speakers successfully coordinate on the use of 'spicy', there will be implications for how the term 'spicy' is used in the menu they are writing for their restaurant. In that case, there will be an eventual, much larger audience they are trying to communicate with, including potential customers, peers in the restaurant business, food critics, etc.

Consider a final case. Two linguists are working on a paper together and disagree on a label for their main argument. One thinks the argument should be called "the argument from context-sensitivity". The other thinks it should be called "the argument from shifting meanings". They argue about this issue both explicitly, and also via metalinguistic negotiation over the relevant terms, but fail to reach a conclusion. After much back and forth, one of the speakers proposes that they postpone the debate about an official label, and use the placeholder label 'the awesome argument' for now, knowing full well that they won't use 'the awesome argument' in the final, published version, or even in any draft they share with colleagues. Imagine that the speakers engage in metalinguistic negotiation on that proposal about 'the awesome argument'. This case resembles the 'spicy' case in that both speakers have as their immediate, intended audience simply their interlocutor. But, unlike in the 'spicy' case, there is no expectation that whatever they coordinate on will eventually carry over to the usage of a term in communication to a broader audience.

Which audience(s) a speaker aims to communicate with in a given metalinguistic negotiation can matter for the dynamics of that dispute. In the Secretariat case, it is plausible that the speakers are more concerned with how what they say comes across to the radio audience than to each other.[29] This dynamic is similar to that of many adversarial political debates. In many such debates, the main goal for the debaters is not to convince their opponents or to achieve coordination on anything. Rather, in many such debates, the primary goal is to make an impact on the audience listening to the debate—for example to rally support for a candidate, a political cause, or a political party. That goal is different from the goal in the 'spicy' case, where there is no outside audience (at least initially), as well as with the 'the awesome argument' case, where the speakers explicitly rule out the idea that their use of the terms will ever extend to a broader audience. Speakers' perceptions of the size and composition of the audience matter for the conversational moves those speakers are likely to make, how they are likely to respond to competing claims, etc.

This issue about intended audience(s) obviously interacts with the question of motivations, discussed in the last section. For example, suppose a speaker is motivated to engage in a metalinguistic negotiation about how to use the term 'athlete' in order to help build political support for reform of how racehorses are treated. It might be that it makes sense for her to argue against an adversary on sports radio—given the potential for her to reach a broader audience—in a way that it wouldn't make sense if she were to consider the prospect of talking to that same person one-on-one, with no one else listening. Or suppose that speakers in a given metalinguistic negotiation use terms in such a way as to put forward particular normative views signaling their group membership or social identity of a certain kind. Again, the speakers' understanding of who is listening to the conversation (or who might potentially be listening) will matter for that motivation and how it is expressed in the conversation.

One final observation about the issue of audience is that we have focused here on the speakers' *intended* audience. But this can come apart from the actual audience that in fact witnesses or is influenced by a metalinguistic negotiation. On one side, there may be eavesdroppers to a metalinguistic negotiation, or a metalinguistic negotiation thought by the participants to be private could be recorded and shared widely. Conversely, a metalinguistic negotiation like the sports radio debate could be intended for a wide audience, but due to unnoticed technical difficulties might end up being observed by a much narrower audience than that. Questions about how the intended audience influences the dynamics of a metalinguistic negotiation are important on their own, as we've explored here. But questions about how metalinguistic negotiations might be interpreted by, or influence, inadvertent audiences, while they go beyond what we can explore in detail here, are also well worth asking.

[29] We emphasize this point in Plunkett and Sundell (2013a).

5 Scope of Normative Claims

The issue of the *scope* of normative claims that are advanced within metalinguistic negotiations is closely related to, but independent of, the issue of audience. When a speaker advocates on behalf of a certain usage of the word 'athlete', or 'spicy', or 'sandwich', what is the range of contexts to which she means her suggestion to apply? Depending on the term, and the speaker's context, beliefs, and motivations, the answer could vary all the way from "every use of the term, in almost any context", to "relevantly similar contexts", to "this particular conversation, and no further". This kind of variation mirrors the fact that *explicit* proposals people make in conceptual ethics vary with respect to whose usage they mean to target.[30] This general point about the variation in the "scope" of the normative claims advanced in metalinguistic negotiations is one that we have emphasized in previous work.[31] And it's been emphasized and explored by others (e.g., Ludlow) working in a similar theoretical space.[32] We discuss it here in part to illustrate important ways in which it interacts with the other dimensions of variation we are exploring.

To see the range in intended scope, consider some examples.

First, consider a realtor working with a young couple who have said they want to live in a "walkable" neighborhood. What counts as "walkable"—the house with parks and shopping a 10-min walk away, across a large, congested road? Or the house that's a 30-min walk from parks and shopping, through an area with short blocks, wide sidewalks, and narrow streets? The agent and the couple engage in metalinguistic negotiation over the meaning of the term 'walkable', arriving at a standard and threshold of "walkability" reflective of the couple's values and desires. But none of them, and certainly not the agent, has any particular expectation that other speakers, even in very similar circumstances, will arrive at the same conclusions.

Second, consider that same realtor, now walking the same neighborhoods with her colleague as part of a tour of neighborhoods they are selling property in. They are discussing how to best advertise those properties on their website, and how the properties' various selling points should be described. They engage in metalinguistic negotiation about how the term 'walkable' should be used in describing different houses, knowing that what they agree on during their walk will impact what they say on their website, how other realtors at the same agencies will describe the neighborhood, and how they themselves will speak on future occasions. But they know that the conclusions they reach won't necessarily apply at other agencies, and that individual agents are likely to modify the meaning in their discussions with individual clients.

Third, consider two city planners reviewing development plans. New developments are required to leave a certain percentage of "usable open space" on any property, but this plan in particular includes the grassy patch behind the dumpster in its calculation of that space, without which the plan would fall short of the requirement.[33] The planners engage in metalinguistic negotiation over the phrase 'usable open space', and 'usable' in particular, arriving at a notion of "usability" that balances the aesthetic and political values of urban design with the practicalities of development. Each speaker has the expectation that in similar conversations between them in the future, the meaning they arrived at will apply. Knowing that their final report will be available to other developers and city employees, they might even expect that other speakers in similar situations in this city will employ the same standard. But they have no expectation that their conclusion would apply at other levels of government, or in any other city.

Finally, consider two political theorists debating the nature of "democracy". What, when it comes down to it, is really required for a society to be "a democracy"? Does it have to do with a kind of broadly egalitarian relationship citizens stand in to each other? The use of voting to decide political outcomes? Or maybe (in addition to one of the things just mentioned) some specific suite of rights that citizens have? The theorists could engage in metalinguistic negotiation with a limited scope—for example, trying to fix usage of the term simply for the sake of their own discussion. But it's also possible to imagine that one or both of them thinks she has the correct view of how the term should be used quite generally for political theorizing, and aims to get a large swath of other theorists to adopt her preferred usage—and even, perhaps, ordinary citizens as well.

Variation in the scope of metalinguistic proposals interacts with both of the previous dimensions of variation we've discussed in a number of ways. To illustrate, if one thinks a wide range of people should change their usage, and one is trying to actually bring about that change, it might make good sense to go in for a metalinguistic negotiation when there is a large audience listening in (as on sports radio) but simply not be worth the time to engage in it one-on-one, especially if the interlocutor is belligerent, intransigent, etc.

[30] For discussion, see Burgess and Plunkett (2013a, b), and Cappelen and Plunkett (2020).

[31] See Plunkett and Sundell (2013a, 2021a), and Plunkett (2015).

[32] See Ludlow (2014).

[33] Thanks to Valerie Friedmann for discussion. This illustration obviously oversimplifies the realities of city planning and development review.

Second, consider the issue of audience. In some cases, the scope of the normative claim advanced in a metalinguistic negotiation will line up with the audience. That is, a speaker might take the very people she perceives to be listening to a metalinguistic negotiation about "X" to be those whom she would like to influence in what they mean by the term 'X'. But this certainly doesn't have to be the case. Things might fail to line up so tidily for all kinds of reasons. To see that, consider the range of explicit views in conceptual ethics one might have and give voice to through the literal content of what one says. For example, someone might think that biologists should change what they mean by the term 'X' and explicitly state that view to a group of people that contains some, or even entirely, non-biologists. Just as that's true of views in conceptual ethics that are communicated explicitly, the same is true for views in conceptual ethics that are communicated via metalinguistic usage.

6 Other Kinds of Normative or Evaluative Claims

Let's turn to the final dimension of variation. This dimension concerns the *kind* of normative and evaluative claims speakers express in metalinguistic negotiation.

It is not necessarily the case that all parties to a metalinguistic negotiation will be using the same normative and evaluative concepts. Nor will they specify given concepts in the same way (e.g., by using the same flavor of deontic modals). For example, certain participants in a metalinguistic negotiation might have views about how, *morally*, people should use a given term, whereas others might be concerned with how to best use that term to promote knowledge, and therefore leave to the side other issues related to an overall moral assessment, or perhaps even reject the idea of moral assessment as such altogether. In some cases, the speakers will be using different normative or evaluative concepts as part of their thinking, but also share, at a more fundamental level, underlying normative or evaluative concepts. For example, they might both care about what a term "really and truly" should mean, and both deploy a shared concept that concerns this kind of "authoritative" ought. But while one speaker happens to think that moral considerations matter for settling this in the context, the other does not. In other instances, it might just turn out that the dispute doesn't express a genuine disagreement after all, since the speakers simply have views about different normative or evaluative topics, picked out by different normative or evaluative concepts.[34]

Variation in what kind of "should" claims someone makes isn't the only kind of variation that is possible here. Consider the capacious nature of "conceptual ethics" as we have presented it. On our understanding, conceptual ethics concerns a cluster of related normative and evaluative issues about thought and talk. Those issues include not only normative ones (e.g., questions about "ought" and "should") but also evaluative ones (e.g., questions about "good" and "bad", or "better" and "worse"). Indeed, given the broad understanding of "conceptual ethics" we are working with, the field should also be seen as involving "aretaic" issues (e.g., questions about how it would be "virtuous" to use a term, and why), which are sometimes treated as separate from "normative" or "evaluative" issues.[35] This raises the question: can speakers engage in "metalinguistic usage" of a term to put forward the full range of views one might have on the topic of conceptual ethics? Or only some of them? Can speakers engage in metalinguistic usage to put forward views about what is a good way to use a term, or views about how it would be virtuous to use a term? Or is the mechanism of metalinguistic usage limited in some way, such that it works as a way to put forward normative views about how a term should be used (perhaps along with closely connected practical views about how to use it), but not these other views in conceptual ethics? These, we think, are questions that need further investigation.

These questions interact closely with the question of what exactly the mechanism of metalinguistic usage really amounts to, and how it works. Is there just one mechanism here, or are there multiple mechanisms that overlap in certain ways? Do different mechanisms fit more easily with the expression of different *kinds* of views in conceptual ethics?

Finally, consider the question of whether speakers involved in metalinguistic negotiations are really making "should" claims at all, or instead making (closely connected, but perhaps distinct) practical claims about *what to do*.[36] If those are distinct kinds of claims, as they are on some metanormative views, then metalinguistic negotiations themselves might vary with respect to whether they involve the expression of "should" claims, or "what to do" claims.[37] Alternatively, it might be that there are two closely connected kinds of metalinguistic dispute, only one of which should count as "metalinguistic negotiation"

[34] Our points above draw from Plunkett (2015).

[35] Our inclusion of these issues is tied to the broad uses of "normative" and "evaluative" that we favor for characterizing conceptual ethics. For connected discussion, see McPherson and Plunkett (2017).

[36] We note this issue in Plunkett and Sundell (2021a). For a view that analyzes "should" judgments in terms of "what to do" judgments, see Gibbard (2003). For views on which these are importantly distinct kinds of judgment, see Hieronymi (2009) and Southwood (2013).

[37] For connected discussion, see Hansen (2019), Khoo and Knobe (2016), and Kocurek et al. (2020).

properly construed, given that the category is understood as involving conflicting normative views. Whether "practical" judgments about "what to do" count as "normative" in the relevant sense is going to be part of what our categorization of things hangs on. Either way, it seems like there are going to be closely connected (though perhaps subtly distinct) disputes here to explore.

This dimension of variation connects back to the previous dimensions in a number of ways. First, take the issue of scope. We discussed that issue in terms of speakers making different "should" claims about which agents (across which contexts) should use a term in a given way, or should use it with a certain meaning. Claims about whether the use of a concept is "good" or "bad"—or how it would be virtuous to use a concept, etc.—can vary in scope in a similar way. For example, we can ask: good or bad as used in which context, by which people? Second, take the issue of audience. In some cases, it might be that a speaker wants to express one kind of normative or evaluative view to one group of people, but not to another—and that carefully navigating the differences between those groups might matter to the uptake or reception of her claims. If one aims to achieve coordination among a group of people, one might be more invested in (at least certain kinds of) "what to do" judgments than if one didn't care about achieving that kind of practical coordination, or if one thought that achieving the relevant kind of practical coordination (with the relevant group of people) was difficult or impossible. Finally, it's plausible that the different kinds of motivations that speakers might have for entering into a metalinguistic negotiation would bear on (or at least should bear on) the normative and evaluative concepts they choose to employ in the metalinguistic negotiation.

7 Objections to Metalinguistic Negotiation, and the "Implementation Challenge" in Conceptual Engineering

We have explored some dimensions of variation in metalinguistic negotiations, and how those types of variation interact. In discussing these issues, we've generally presented things in terms of dimensions of variation amongst metalinguistic negotiations that actually exist. And we do believe that many of the kinds of cases we've discussed are in fact realized. But there could be varieties of metalinguistic negotiation that are possible in some philosophically useful sense, but which we rarely or never see actualized. The questions of "How wide is the full range of potential metalinguistic negotiations?" and "How wide is the range of metalinguistic negotiations that actually occur?" are separate, though both are worth exploring.

We consider it likely that much of the variety we've discussed here is actually instantiated. This seems especially likely with regard to the matters of motivation, audience, and scope. But we are less sure about the full range of normative and evaluative claims we discussed in the last section. Are there indeed metalinguistic negotiations involving (in the first instance) evaluative views, rather than normative ones? It's simply not obvious one way or the other whether there are actual conversations best analyzed in this way. Our hope is that future work can examine which of the hypothetical kinds of disputes we've discussed are "live options" in this sense, along with further questions about the hypothetical options, and additional dimensions of variation. In putting these issues on the table, we hope this paper can pave the way for fruitful new exploration of metalinguistic negotiation and related phenomena.

In closing, we note two final points.

First, recall that, at the start of this paper we claimed our discussion could help us respond to certain critiques of our account that one might otherwise be drawn to. The reason is that a number of existing criticisms of our framework take us to have a narrower understanding of what's involved in metalinguistic negotiation than we actually do. Matthew Shields, for example, claims that our discussion of a metalinguistic dispute involving 'torture' problematically assumes the participants to be good faith actors trying to form the correct view about which concept this word should be used to express in the context at hand.[38] This claim in turn forms the basis for a critique of our more general view as overly narrow and as excluding important cases where a metalinguistic-negotiation-like phenomenon occurs, but where parties to the dispute are in different positions of power, and may or may not be acting in good faith.

We think the cases Shields draws attention to are indeed significant, well analyzed as involving something like metalinguistic negotiation, and very much worth exploring. But there is no assumption of good faith, or of a specific focus on inquiry as the speakers' motivation, in our original discussion of the 'torture' case specifically, much less metalinguistic negotiation more generally.[39] As we have emphasized, our account makes room for speakers having a range of motivations, including ones where successfully forming views about a topic is either subservient to some broader practical goal, or even totally beside the point. And nothing in the original account requires that speakers act in good faith, or sincerely believe the claims they metalinguistically advance. In short, these cases aren't problems for our account of metalinguistic negotiation or of our assumptions about core cases, but rather interesting and important instances of the very phenomenon that we aim to draw attention to.[40]

[38] Shields (2021).

[39] See Plunkett and Sundell (2013a, b).

[40] This point matters not just for potential critiques of our views, such as Shields's, but also just for understanding how our framework relates to other discussion of cases of metalinguistic disputes that dif-

Second, recall that we suggested our discussion of variation in metalinguistic negotiations could matter for other issues in philosophy. We want to float one place where this might be so, namely in connection to issues about the "implementation" of conceptual engineering proposals. As we understand it, "conceptual engineering" combines the normative and evaluative work of conceptual ethics with work on conceptual innovation, and with the implementation of proposed innovations.[41] Roughly speaking, the *innovation* component of such projects involves determining how our concepts, words, etc. might be improved by doing things like revising their meaning, replacing them with alternative representations, supplementing them with new words or concepts, or eliminating them altogether. The *implementation* component involves (again, speaking roughly) trying to actually bring about those changes in practice. A key question about the implementation element of such projects is how—or even whether—making these kinds of changes to our words or concepts can be accomplished.

With that general question in mind, we can ask how helpful metalinguistic negotiation might be in implementing conceptual engineering proposals. We think that the potential for metalinguistic negotiation in implementing conceptual engineering projects has been underexplored, and sometimes underappreciated. This is for two reasons. First, much of the discussion of the "implementation challenge" for conceptual engineering has focused on a particular subset of conceptual engineering proposals: namely, those aimed at effecting large-scale linguistic changes, such as changes to the meanings of words in a public language.[42] Second, and conversely, the most commonly discussed examples of metalinguistic negotiation are often characterized as situations where a meaning is being advocated only to a specific speaker, or for purposes of a specific conversation. In other words, discussion of conceptual engineering has tended to focus on large-scale proposals, while discussion of metalinguistic negotiation has tended to focus on negotiations with small-scale audiences or ambitions. Put together, these two tendencies would make it easy to think that metalinguistic negotiation is too local or small-scale a phenomenon to have much bearing on the "implementation challenge" for conceptual engineering.[43] But this would be a mistake.

To see why, let's start with the issue of how conceptual engineering proposals can vary from each other. Consider again the three key components of conceptual engineering projects, as we understand them: "conceptual ethics", "conceptual innovation", and "conceptual implementation". The nature of the "implementation challenge" in a particular case will inevitably depend on *what* innovation is being implemented, and on the kinds of views in conceptual ethics that motivate the proposed change. As we emphasize in Sect. 2, views in conceptual ethics themselves vary both with respect to the kinds of object they concern—words, concepts, patterns of usage, etc.—and also with respect to the scope of the normative or evaluative claims they involve concerning those objects.

For example, a conceptual ethics view *loosely* described as the belief that "the word 'sandwich' should include hot dogs" could, when made more precise, amount to the view that the conventionally encoded content of 'sandwich' in the public language of some large-scale speech community should be changed in such a way as to always express a concept that has hot dogs in its extension. That would be a large-scale proposal indeed. But that loose description could also reflect a very different kind of view in conceptual ethics: say, the view that a select group of speakers in some very local context, should talk, for certain purposes, as if hot dogs are "sandwiches". These views vary both with respect to scope (e.g., large vs. small populations, or all contexts vs. specific sets of contexts) and also with respect to the objects the views concern (e.g., meanings in a public language vs. say, speaker meaning, or use). Both beliefs are views in conceptual ethics. Both beliefs could motivate attempts at implementation. But the implementation challenges for those proposals would look very different.

The upshot is that variation in the nature of conceptual engineering views and proposals will interact with how useful metalinguistic negotiation will be for helping to bring

Footnote 40 (continued)

fer from the cases we have focused on. For example, we think that in light of the discussion in this essay, our framework can incorporate key cases discussed by Podosky (2022) and Davies (2021).

[41] For further discussion of this understanding of conceptual engineering and its connections to conceptual ethics, see Burgess and Plunkett (2020) and Cappelen and Plunkett (2020). Note that there are subtle differences between these accounts. However, those differences won't matter to us here. For a range of other accounts of what "conceptual engineering" is, see the papers collected in Burgess et al. (2020).

[42] For a view that emphasizes these larger kind of ambitions for key projects in conceptual engineering within philosophy (including his reading of certain paradigmatic examples of such projects), see Cappelen (2018). For arguments that push back against the idea that important conceptual engineering projects (including ones that are often taken to be paradigmatic examples of conceptual engineering within philosophy, such as Haslanger (2000)) really have this ambition, see Pinder (2021) and Riggs (2019).

[43] Koslow (2022) discusses Kevin Scharp's proposals regarding truth (in Scharp (2013)) as a *limiting* case with respect to the small size of the linguistic population targeted, despite Scharp's proposal still being directed at a group. (Namely, logicians, or those facing "liar-shaped predicaments".) If Scharp's proposal—directed at a small group, but a group nonetheless—is a limiting case, then changes directed at individual speakers or specific conversations—like the changes advocated in classic cases of metalinguistic negotiation—are unlikely to even register as cases of conceptual engineering.

about the relevant changes. If one is, say, trying to bring about small changes in the patterns of usage tied to the phrase 'pragmatic encroachment', as used by certain people (a particular group of philosophers) in certain contexts (when discussing a certain collection of papers during a single epistemology conference in 2022), then engaging in small-scale, local metalinguistic negotiations (say, with that very group of people, during that conference) is potentially a worthwhile strategy. If one is trying to change the meaning of a term like 'democracy', as used by everyone in Portugal, across all contexts where they use that term, then perhaps less so.[44]

So one reason metalinguistic negotiation could matter when it comes to the implementation of proposals in conceptual engineering is that those proposals sometimes relate to the meanings (or even just the use) of terms as they are employed by smaller groups of speakers, or with respect to narrowly constrained sets of contexts. These are exactly the kinds of changes that even the original, most widely discussed examples of metalinguistic negotiation are well suited to bring about.

If those were the only kinds of metalinguistic negotiations, it would be easy to think the discussion ends there. That is, one might think that the mechanism of metalinguistic negotiation will only be effective for helping to bring about a certain subset of conceptual engineering proposals—those tied to smaller-scale changes. To be fair, that subset of conceptual engineering proposals might still cover a large number of important examples of real-life conceptual engineering, including cases like the house-hunting couple's determination of a meaning for "walkable", or the restaurant owners' determination of a meaning for "spicy". It might also cover some important examples in philosophy, including some of the more widely cited illustrations of philosophical conceptual engineering. For example, Kevin Scharp's work on 'true', which aims to bring about changes about how a certain limited group of theorists use terminology tied to "truth"-talk, in certain contexts. Or certain cases of Carnap-style explication, in which the meaning of philosophical terms is settled for purposes of philosophical or scientific discussion, with no intention of exporting those changes to broader, everyday discourse.[45]

But is there good reason to stop with this subset of conceptual engineering proposals? We think not. As we've argued in this paper, metalinguistic negotiations do not necessarily express views in conceptual ethics that concern local or small-scale contexts only. And participants in a metalinguistic negotiation do not necessarily have a small audience in mind when they express those views. Participants in the "athlete" debate, for example, might not care much at all about how their claims land with, or influence, their immediate conversational partners. Rather, they may be expressing a view about how 'athlete' should be used, across a broad range of contexts, for a large-scale speech community, and they may intend to express those views to a large-scale audience of radio listeners. These kinds of variation—in motivation, audience, scope, and perhaps even in the kind of normative or evaluative claim—open the door to metalinguistic negotiations playing a role in even large-scale proposals in conceptual engineering.

The importance of such variation for implementation issues in conceptual engineering has, we think, been underappreciated. To illustrate, consider Rachel Sterken's discussion of metalinguistic negotiations in her paper "Linguistic Interventions and Transformative Communicative Disruption".[46] She writes that "metalinguistic negotiations are limited in their scope—the aim is to settle what a given word should mean *in the context of a given communicative exchange*."[47] She therefore distinguishes metalinguistic negotiations—where speakers "needn't have diachronic intentions to change the meaning for the linguistic community as a whole"—from what she labels *linguistic interventions*, which are, by definition, more ambitious in this sense.[48]

While it's true that metalinguistic negotiations do not necessarily involve this ambition, part of our aim here has been to emphasize that they can, and, likely often, do involve broader ambitions. This suggests that "metalinguistic negotiation" and "linguistic interventions" in Sterken's sense are not mutually exclusive, but rather closely related ideas with significant overlap. Sterken defines "linguistic interventions" as "communicative activities on the part of a speaker that (intentionally and strategically) attempt to change the word-meaning pairs in circulation".[49] This definition is broad enough to include conversations that do not take the form of a dispute. So not all linguistic interventions are metalinguistic negotiations. And since speakers can engage in metalinguistic negotiations without being aware they are doing so, the attempts to influence the meaning or use of a term via metalinguistic negotiation are not necessarily intentional or strategic. So not all metalinguistic negotiations are linguistic interventions in Sterken's sense. Nevertheless, once our observations here are taken into account, Sterken's definition

[44] For further discussion of these (and other) kinds of variation here, see Burgess and Plunkett (2013a, b), and Cappelen and Plunkett (2020).

[45] See Scharp (2013) and Carnap (1947/1956).

[46] Sterken (2020).

[47] Sterken (2020, p. 420, emphasis Sterken's).

[48] Sterken (2020, p. 420).

[49] Sterken (2020, p. 418).

should allow that metalinguistic negotiation is one common form that linguistic interventions can take.

Other discussions of implementation in conceptual engineering make similar assumptions about the small-scale scope of metalinguistic negotiations, leading philosophers to either explicitly or implicitly reject metalinguistic negotiation as a potential mechanism for broad-based linguistic changes. Sigurd Jorem, for example, mentions metalinguistic negotiation as a possible account of the speech acts involved in some cases of conceptual engineering.[50] But he takes the feasible modes of implementation for conceptual engineering to involve changes to speaker meaning, to word meaning for purposes of a specific conversation, or to new or artificial variants on a language. When it comes to changing the standing meaning of terms in a public language, Jorem is more pessimistic that the implementation problem can be addressed. On Jorem's view, if there is a role for metalinguistic negotiation for such semantic changes, it would *not* be because individual negotiations can involve broader ambitions, but rather because the kinds of changes that are actually feasible are themselves at mostly smaller scales.

In contrast to Jorem, Matthieu Queloz and Friedemann Bieber treat intentional, large-scale semantic change as a possibility. (Though they argue that it is of dubious desirability.)[51] Despite their thinking it is possible, they discuss the possibility of large-scale change only in terms of institutional mechanisms—changes like those dictated by the World Health Organization in categorizing diseases, for example. This focus is appropriate enough given their specific dialectical goals, but again reflects the tendency to discuss large-scale semantic changes exclusively in terms of large-scale mechanisms. Allison Koslow is an exception to that pattern. She emphasizes the role of the accretion of small interventions in propagating large-scale semantic changes.[52] But as her focus is on descriptive patterns in natural language change that could make certain conceptual engineering projects more or less likely to succeed, she does not discuss specific mechanisms of intentional change at either the large or small scale.

We don't take it to be a fatal flaw in any of the work mentioned here that it doesn't address the subject of metalinguistic negotiation as a mechanism for more ambitious projects in conceptual engineering. That is because the arguments of these authors are all primarily focused on other aspects of the relevant questions.[53] What we suggest here is rather that attention to metalinguistic negotiation simply has not played a significant role in discussions of implementation when it comes to more ambitious projects in conceptual engineering aimed at large-scale linguistic changes, such as changes in the semantics of words in a public language. We suggest this is not a surprising result given the perceived focus on small-scale motivations, audiences, and claims in many of most widely discussed examples of metalinguistic negotiation, and the perception that these examples represent something like the full range of possible metalinguistic negotiations.

When the full range of cases of metalinguistic negotiation is accurately appreciated, things look different. Moves in a metalinguistic negotiation can target a wide audience—wider than just the other parties who happen to be present. They can communicate claims with a wide scope—targeting the meaning or usage of wide swaths of speakers or large-scale speech communities. They can be motivated by an interest in affecting the meaning or usage of a term over a wide range of contexts—much wider than simply the specific conversation in question. Even the possible variation in *type* of normative claims involved in a metalinguistic negotiation might be relevant here, in that it more fully reflects the range of claims found across different projects in conceptual engineering. Metalinguistic negotiations might start to look like increasingly plausible loci for efforts to create broader linguistic change, such as changes in the semantics of words in a public language.

Even metalinguistic negotiations where the motivations, intended audience, and scope of claims reflect only very local interests *could* prove to be important in addressing implementation challenges. This is for two reasons. First, as noted above, some conceptual engineering proposals explicitly take on narrower targets. Second, even if a conceptual engineering proposal involves something more ambitious, small-scale interventions might well have wider ramifications. Once the full range of metalinguistic negotiations is taken into account, we suspect the phenomenon could have an even bigger role to play in addressing the real-world possibilities for various projects in conceptual engineering, including more ambitious projects in conceptual engineering aimed at large-scale linguistic changes, such as changes in the semantics of words in a public language. Investigating that role in detail is more than we can do here, but we hope to have made a *prima facie* case for the worthiness of further exploration. If metalinguistic negotiation is a significant feature of linguistic communication quite generally, as we think it is, then understanding the range of variation in its

[50] Jorem (2021, p. 190).

[51] Queloz and Bieber (2021).

[52] Koslow (2022).

[53] This might not be so, however, for Cappelen (2018)'s discussion of metalinguistic negotiation and conceptual engineering. Even there, however, we can take his comments to reflect a broader skepticism about metalinguistic negotiation, and a broader pessimism about

Footnote 53 (continued)

intentional semantic revision, rather than being focused on the view that metalinguistic negotiations are too limited in scope to bear on projects in conceptual engineering.

instances is independently worthwhile. But we also think that connections to issues concerning the implementation of conceptual engineering proposals, which we have discussed briefly here, reflect the fact that a better understanding of metalinguistic negotiation is likely to have implications for a range of philosophical debates.

Acknowledgements Thanks to Pedro Abreu, Chris Barker, Alexis Burgess, Elisabeth Camp, Herman Cappelen, Andy Egan, Jesse Ferraioli, Valerie Friedmann, Sean Kim, Devontae Lacasse, Zachary Lang, Peter Ludlow, Tristram McPherson, Mark Richard, Adrian Russian, Rachel Sterken, Giulia Terzian, Zoe Thierfelder, Amie Thomasson, Fangzhou Yu, anonymous referees, and audiences at the Nova University of Lisbon and the University of Warsaw for helpful discussion about ideas in this paper or for feedback on previous drafts of this paper.

Funding Funding information is not applicable/No funding was received.

Data availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

Abreu P (2023) Metalinguistic negotiation, speaker error, and charity. Topoi. https://doi.org/10.1007/s11245-023-09910-9
Barker C (2002) The dynamics of vagueness. Linguist Philos 25:1–36
Barker C (2013) Negotiating taste. Inquiry 56(2–3):240–257
Belleri D (2017) Verbalism and metalinguistic negotiation in ontological disputes. Philos Stud 174(9):2211–2226
Bolinger RJ (2022) Metalinguistic negotiations in moral disagreement. Inquiry 65(3):352–380
Burgess A, Plunkett D (2013a) Conceptual ethics I. Philos Compass 8(12):1091–1101
Burgess A, Plunkett D (2013b) Conceptual ethics II. Philos Compass 8(12):1102–1110
Burgess A, Plunkett D (2020) On the relation between conceptual engineering and conceptual ethics. Ratio 33(4):281–294
Burgess A, Cappelen H, Plunkett D (2020) Conceptual engineering and conceptual ethics. Oxford University Press, Oxford
Cappelen H (2018) Fixing language: an essay on conceptual engineering. Oxford University Press, Oxford
Cappelen H, Plunkett D (2020) A guided tour of conceptual engineering and conceptual ethics. In: Burgess A, Cappelen H, Plunkett D (eds) Conceptual engineering and conceptual ethics. Oxford University Press, Oxford, pp 1–26
Carnap R (1947/1956) Meaning and necessity: a study in semantics and modal logic. University of Chicago Press, Chicago
Davies A (2021) Identity display: another motive for metalinguistic disagreement. Inquiry 64(8):861–882
Dworkin R (1986) Law's empire. Belknap Press, Cambridge
Eidelson B (2023) The etiquette of equality. Philos Public Aff 51(2):97–139
Gibbard A (2003) Thinking how to live. Harvard University Press, Cambridge
Hansen N (2019) Metalinguistic proposals. Inquiry 1–2:1–19
Haslanger S (2000) Gender and race: (what) are they? (What) do we want them to be? Noûs 34(1):31–55
Hieronymi P (2009) The will as reason. Philos Perspect 23(1):201–220
Horgan T, Timmons M (1993) New wave moral realism meets moral twin Earth. In: Heil J (ed) Rationality, morality, and self-interest. Rowan and Littlefield Publishers, Lanham, pp 115–133
Jorem S (2021) Conceptual engineering and the implementation problem. Inquiry 64(1–2):186–211
Khoo J (2017) Code words in political discourse. Philos Top 45(2):33–64
Khoo J (2020) Quasi indexicals. Philos Phenomenol Res 100(1):26–53
Khoo J, Knobe J (2016) Moral disagreement and moral semantics. Noûs 50(2):109–143
Kocurek AW, Jerzak E, Rudolph RE (2020) Against conventional wisdom. Philos Imprint 20(22):1–27
Koslow A (2022) Meaning change and changing meaning. Synthese 200(2):1–26
Ludlow P (2008) Cheap contextualism. Philos Issues 18:104–129
Ludlow P (2014) Living words: meaning undetermination and the dynamic lexicon. Oxford University Press, Oxford
Mankowitz P (2021) How to have a metalinguistic dispute. Synthese 199(3–4):5603–5622
McPherson T, Plunkett D (2017) The nature and explanatory ambitions of metaethics. In: McPherson T, Plunkett D (eds) The Routledge handbook of metaethics. Routledge, New York, pp 1–25
Nunberg G (2018) The social life of slurs. In: Fogal D, Harris D, Moss M (eds) New work on speech acts. Oxford University Press, Oxford
Odrowąż-Sypniewska J (2023) Spicy, tall, and metalinguistic negotiations. Topoi. https://doi.org/10.1007/s11245-023-09930-5
Pinder M (2021) Conceptual engineering, metasemantic externalism and speaker-meaning. Mind 130(517):141–163
Plunkett D (2015) Which concepts should we use? Metalinguistic negotiations and the methodology of philosophy. Inquiry 58(7–8):828–874
Plunkett D (2016) Negotiating the meaning of "law": The metalinguistic dimension of the dispute over legal positivism. Leg Theory 22(3–4):205–275
Plunkett D, Sundell T (2013a) Disagreement and the semantics of normative and evaluative terms. Philosophers' Imprint 13(23):1–37
Plunkett D, Sundell T (2013b) Dworkin's interpretivism and the pragmatics of legal disputes. Leg Theory 19(3):242–281
Plunkett D, Sundell T (2014) Antipositivist arguments from legal thought and talk: the metalinguistic response. In: Hubb G, Lind D (eds) Pragmatism, law, and language. Routledge, New York, pp 56–75
Plunkett D, Sundell T (2021) Metalinguistic negotiation and speaker error. Inquiry 64(1–2):142–167
Plunkett D, Sundell T (2021) Metalinguistic negotiation and matters of language: a response to cappelen. Inquiry. https://doi.org/10.1080/0020174X.2021.1983456
Plunkett D, Sterken RK, Sundell T (2023) Generics and metalinguistic negotiation. Synthese 201(50):1–46
Podosky P-M (2022) Agency, power, and injustice in metalinguistic disagreement. Philos Quart 72(2):1–24
Quaranto A (2022) Dog whistles, covertly coded speech, and the practices that enable them. Synthese 200(4):1–34
Queloz M, Bieber F (2021) Conceptual engineering and the politics of implementation. Pac Philos Q 3:670–691
Riggs J (2019) Conceptual engineers shouldn't worry about semantic externalism. Inquiry. https://doi.org/10.1080/0020174X.2019.1675534
Saul J (2018) Dogwhistles, political manipulation, and philosophy of language. In: Fogal D, Harris DW, Moss M (eds) New work on speech acts. Oxford University Press, Oxford, pp 360–383
Scharp K (2013) Replacing truth. Oxford University Press, Oxford
Shields M (2021) Conceptual domination. Synthese 199(5–6):15043–15067

Soria-Ruiz A (2023) Metalinguistic usage is not conversational implicature. Topoi

Southwood N (2013) "The thing to do" implies "can". Noûs 47(3):61–72

Stanley J (2015) How propaganda works. Princeton University Press, Princeton

Sterken RK (2020) Linguistic interventions and transformative communicative disruption. In: Cappelen H (ed) Conceptual engineering and conceptual ethics. Oxford University Press, Oxford, pp 417–434

Sundell T (2011a) Disagreement, error, and an alternative to reference magnetism. Austral J Philos 90(4):743–759

Sundell T (2011b) Disagreements about taste. Philos Stud 155(2):267–288

Sundell T (2016) The tasty, the bold, and the beautiful. Inquiry 59(6):793–818

Sundell T (2017) Aesthetic negotiation. In: Young J (ed) The semantics of aesthetic judgment. Oxford University Press, New York City, pp 82–105

Thomasson AL (2016) Metaphysical disputes and metalinguistic negotiation. Anal Philos 57(4):1–28

Thomasson AL (2020) A pragmatic method for normative conceptual work. In: Burgess A, Cappelen H, Plunkett D (eds) Conceptual engineering and conceptual ethics. Oxford University Press, Oxford, pp 435–458