

# The Limits of Metalinguistic Negotiation: The Role of Shared Meanings in Normative Debate

François Schroeter<sup>1,\*</sup>; Laura Schroeter<sup>1</sup>; Kevin Toh<sup>2</sup>

<sup>1</sup> Philosophy discipline, University of Melbourne, Victoria, Australia

<sup>2</sup> Faculty of Laws, University College London, London, United Kingdom

\* Corresponding author. Email: [fschro@unimelb.edu.au](mailto:fschro@unimelb.edu.au)

## Abstract

According to philosophical orthodoxy, the parties to moral or legal disputes genuinely disagree only if their use of key normative terms in the dispute express the same meaning. Recently, however, this orthodoxy has been challenged. According to an influential alternative view, genuine moral and legal disagreements should be understood as metalinguistic negotiations over which meaning a given term should have. In this paper, we argue that the shared meaning view is motivated by much deeper considerations than its recent critics recognize, and that much would be lost in opting for the explanation of normative disputes as metalinguistic negotiations.

**Keywords:** Disagreement; metaethics; legal philosophy; shared meanings; metalinguistic negotiation

---

The nature of fundamental disagreement in the moral and legal domains has come to be a central issue in contemporary metaethics and legal philosophy.<sup>1</sup> The standard view is that the parties to a moral or legal dispute genuinely disagree only if they use normative terms with same meanings. Otherwise, the two parties are simply speaking past each other, making claims about different topics. To explain disagreement, realists see their task as that of working out how different speakers can pick out the very same property despite their divergent views about its instantiation conditions. In contrast, expressivists and truth-relativists argue that even proponents of wildly varying normative positions can have genuine normative disagreements when they share the nondescriptive or relativist meanings of normative terms. Others, including some contextualists, reject the idea that fundamental normative disagreement is possible when discussants are very far apart in their normative views, and seek to explain away the appearance of genuine disagreement in such cases. In sum, despite its range and diversity, the philosophical debate about how to explain fundamental disagreement has been structured by a shared assumption: in order to engage in a genuine normative disagreement, speakers *must share the same meanings* of key moral and legal terms.

This shared assumption has been forcefully challenged in an influential series of papers by

David Plunkett and Tim Sundell (hereinafter, ‘P&S’) (2013a, 2013b, 2021). Drawing on work in linguistics, P&S argue that fundamental moral and legal disagreements need not involve shared meanings since the substantive point of contention can be communicated through purely pragmatic mechanisms. In particular, they suggest that fundamental normative disagreements may be *metalinguistic negotiations* (MLNs), in which speakers jockey over which meanings key moral or legal terms *should* have.<sup>2</sup> If correct, this account of disagreement as MLN would significantly simplify the task of explaining the meaning of normative terms, and undercut much of the contemporary debate about fundamental disagreement. This analysis of disagreement as MLN has also been highly influential in the literature on conceptual engineering (e.g., Cappelen 2018; Burgess, Cappelen, and Plunkett 2020), where it has been extended to explain disagreements in a wide range of nonnormative domains such as metaphysics, logic, and social kinds (e.g., Thomasson 2017; Sterken 2020; Kouri Kissel 2021).

In this paper, we present the case against MLN and in favor of standard shared meaning accounts of normative disagreement. P&S are right to point out that the shared meaning hypothesis has not been systematically defended in the existing literature, so the mere possibility of a purely pragmatic explanation of normative disagreement, such as MLN, highlights the need for a such a defence.<sup>3</sup> The main aim of this paper is to answer this challenge. We argue that shared meanings provide a better explanation of the epistemic and coordinating commitments of the parties to normative disagreements than a pragmatic account such as MLN. In particular, shared meanings play a crucial role in explaining the dynamics of normative debate: they are the fulcrum on which normative debates turn. Our conclusion is that charitable interpretation of normative and legal terms favors assigning shared meanings, rather than divergent meanings, even in cases of deep disagreement.

The paper proceeds as follows. Section 1 outlines the argument for MLN as a possible explanation of genuine moral and legal disagreements. In section 2, we challenge the claim that MLN can explain genuine normative disagreement. In section 3, we highlight the epistemic and coordinating commitments speakers normally undertake in moral and legal debate. Then, in section 4, we argue that these commitments support, as the default interpretation, the standard view that speakers share the same meaning in moral and legal disagreement. Finally, in section 5, we take a parting look at MLN and suggest that it is ill-suited as a vehicle for serious debate.

## 1. Metalinguistic disputes and genuine disagreement

Consider the following argument form that P&S take to be implicit in many metaethical and jurisprudential debates (2013b, 247–48, 252–53; 2013a, 5–8). The simplest form of the argument is a move from the intuition that two parties to a dispute involving the application of a term like ‘morally right’ or ‘legal’ to a particular case are not talking past each other; they are engaging in a genuine disagreement. The conclusion is that the parties use the words ‘morally right’ or ‘legal’ with the same meaning. P&S argue that this move is too quick since there is another possible explanation of the disagreement that doesn’t posit sameness of meaning.

Here is how P&S break down the reasoning from disagreement to sameness of meaning:

1. **Genuine disagreement:** Individuals genuinely disagree when they engage in linguistic disputes in which the parties affirm or deny a given sentence involving normative terms like ‘morally right’ or ‘legal,’ even if the individuals’ settled dispositions to apply these terms to cases diverge.
2. **Disagreement is psychological:** Genuine disagreement involves individuals’ acceptance of rationally incompatible thought contents (e.g., beliefs with incompatible truth-conditions or moral/legal endorsements of incompatible actions).
3. **Sentential meaning:** The parties to moral/legal disputes literally express rationally incompatible sentential meanings when they affirm or deny sentences using these terms (e.g., their sentences express incompatible truth conditions or incompatible moral/legal attitudes).
4. **Word meaning:** In these disputes, the parties’ use of these key terms literally express the same concepts (e.g., the same descriptive application-conditions or same expressive role).

Step 1 is supported by intuition: the dispute in question is not a case of two individuals talking past each other or engaging in a purely verbal dispute about which linguistic labels to use. Genuine disagreements involve substantive questions distinctive of the moral or legal domain. Step 2 clarifies the nature of disagreement: disagreement is not an activity but a relation between the thought contents of the parties in the dispute. In order to disagree, the parties must have rationally incompatible thought contents.<sup>4</sup> The transition to step 3 is the key step from P&S’s point of view: from the claim that incompatible thought contents are expressed in a dispute, we conclude that those thought contents are *literally expressed* as the semantic contents of the sentences used in that dispute.<sup>5</sup> The transition from step 3 to step 4 then moves from sentential meaning to word meaning: if both parties to the dispute use a given sentence to literally express the same sentential content, then (probably) the words they used in that sentence express the same contents—i.e., concepts. P&S don’t challenge that transition.

Once again, step 3, the transition from a genuine psychological disagreement to the literal semantic content of the sentence used in the dispute, is P&S’s main target. Their key claim is that there are ways of communicating a psychological disagreement via purely pragmatic means *without* positing shared literal contents of the words used to express those disagreements—e.g., as we do with some conversational implicatures. So we shouldn’t jump to posit shared meanings unless we can rule out alternative pragmatic accounts (2013a, 5, 25).

P&S suggest that many moral and legal disputes can be explained as cases of MLN (2013a, 3; 2013b, 248). On this view, the parties to a legal dispute disagree about the metalinguistic question of which concept (application conditions, expressive role) their shared term, like ‘defamation’ or ‘culpable,’ *should* have. And the point of the conversation is that each party is seeking to persuade the other to adopt her preferred concept.

One might worry that this account reduces moral and legal disagreements to mere verbal disputes—i.e., arguments over the question of how to pair sounds with meanings. P&S *do* think that a verbal dispute is a genuine psychological disagreement. In it, the parties have incompatible commitments about whether a syntactically individuated item, say ‘culpable,’ should be applied to particular cases (2013a, 15). P&S also think that the parties may have a second, more substantive

disagreement, which explains our interest in moral and legal disputes, and also why these disputes cannot be resolved just by disambiguation. This substantive disagreement concerns which concept (intension, expressive role) should be adopted to fill the *functional role interlocutors associate with the relevant word in their conversational context*. This second type of disagreement is what vindicates the intuition that there is a *nontrivial* disagreement expressed by moral and legal disputes (2013a, 19–24).

To get a clearer picture of P&S’s account of nontrivial legal or moral disagreement, consider their ‘torture’ example (2013a, 19–20). Two individuals have different criteria for applying ‘torture’—one accepts the UN definition, the other accepts the alternative definition pushed by the Bush administration. During a policy debate, their different commitments could lead to an exchange like the following:

**Anne:** Waterboarding is torture.

**Ben:** aterboarding is not torture.

According to P&S, this would not be a trivial verbal dispute:

After all, in the content of discussion about the moral or legal issues surrounding the treatment of prisoners, there is a substantive question about which definition is *better*. By employing the word ‘torture’ in a way that excludes waterboarding, [Ben] communicates (though not via literal expression) the view that such a usage is appropriate to those moral or legal discussions. [...]

More specifically, it has something to do with *which concept should play a functional role* that concerns matters of how we navigate our decisions about how to treat others, what to hold each other responsible for doing, and how to live more generally. (2013a, 19–20)

P&S’s suggestion, then, is that the two parties pragmatically conduct a substantive metalinguistic disagreement about ‘which concept is best suited to play a certain functional role in thought and practice’ (2013a, 21).

Moreover, this sort of metalinguistic disagreements is not confined to cases of actual communication. According to P&S, individuals can have substantive metalinguistic disagreements about moral questions even if they have never communicated with each other and even if they do not share the same language:

[I]t is entirely sensible to suppose that *before* they engage in any linguistic exchange—before they are even aware of one another’s existence, much less one another’s language—[two speakers could] have views about which concept should play this important functional role in thought and practice. They [could] have those views entirely independently of any awareness of the other person, or of the existence of another language. (2013a, 21)

Substantive metalinguistic disagreement, then, is not about a particular linguistic expression per se, it’s about which concepts (e.g., intensions, expressive roles) should fulfill a specific

functional role. The psychological commitment to specific functional roles that concepts are to fill is what explains genuine substantive disagreement at the psychological level.

## 2. Vindicating genuine disagreement?

In this section, we examine P&S's claim that MLN provides a possible explanation of substantive disagreements in legal and moral cases. We argue that P&S's account of substantive metalinguistic disagreement seems vulnerable to precisely the same worry that they raise against interpretations of linguistic disputes that posit shared meanings. On the face of it, MLN doesn't seem to vindicate genuine substantive disagreement. Our conclusion is that P&S need to further explain the sense in which speakers have genuine disagreements when they engage in MLN.<sup>6</sup>

Let's start with P&S's account of what's required for genuine disagreement:

*Disagreement Requires Conflict in Content (DRCC)*: If two subjects A and B disagree with each other, then there are some objects p and q (propositions, plans, etc.) such that A accepts p and B accepts q, and p is such that the demands placed on a subject in virtue of accepting it are rationally incompatible with the demands placed on a subject in virtue of accepting q. (Perhaps, though not necessarily, in virtue of q entailing not-p.) (2013a, 11)

Most metaethicists and legal philosophers would be happy to accept this conception of genuine disagreement. The notion of rational incompatibility of thought contents seems an apt and theoretically neutral way of characterizing the logical or quasilogical relations they seek to explain. Both metaethicists and legal philosophers seek interpretations of key moral and legal terms that can vindicate the logical validity of deductive inferences, as can be seen in expressivists' various attempts to address the so-called Frege-Geach problem about the meanings of normative terms in embedded and unembedded contexts.<sup>7</sup> So it's natural to assume that, in the context of metaethics and legal philosophy, we should interpret *rational incompatibility* among thought contents posited by DRCC as having the modal and normative properties of logical inconsistency. On this reading, two beliefs are rationally incompatible only if they have strictly inconsistent truth-conditions: i.e., there is no possible world that would make both beliefs true. And two plans are rationally incompatible only if there is no possible world in which both plans are executed. But logical inconsistency requires more than this strict modal inconsistency: e.g., 'water is potable' is not logically inconsistent with 'H<sub>2</sub>O is not potable' even if 'water is potable' and 'H<sub>2</sub>O is potable' have the same truth-conditions. Intuitively, what's distinctive of logical relations is that they are knowable in a distinctive way—i.e., in virtue of their logical or semantic form rather than in virtue of their substantive semantic content. We'll assume provisionally that genuine moral and legal disagreement requires logical inconsistency. But we will return to this issue in the final paragraph of this section.

As many metaethicists and legal philosophers have noted, it is not obvious that participants in moral and legal debates are in strict disagreement if their understanding of the relevant moral and legal terms diverges systematically. Consider again the example of 'torture.' P&S worry that positing *shared literal contents* for Anne's and Ben's uses of the term 'torture' conflicts with an important prima facie interpretive constraint—a constraint which requires us to assign semantic contents in such a way as to respect an individual speaker's dispositions to use that term:

Regardless of one's precise views in semantics, it should be uncontroversial that at least one crucial type of data for figuring out what a speaker means by a term T are facts about the speaker's usage of T—patterns of usage that reflect her disposition to apply that term one way or another, more generally. (2013a, 16)

Of course, Anne and Ben probably have a great deal of overlap in the patterns of understanding they associate with 'torture': they may have similar beliefs about its effects, and negative attitudes towards the actions they classify as 'torture,' and their dispositions to classify cases may overlap for most cases. But their dispositions to apply the word 'torture' systematically diverge in the case of waterboarding. According to P&S, this divergence in application dispositions is *prima facie* evidence that Anne and Ben associate different extensions, and therefore different meanings or concepts, with the term 'torture' (2013a, 16). But then Anne and Ben are *not* strictly disagreeing at the level of literal content: the literal content of Anne's sentence is *that waterboarding is torture<sub>UN</sub>*, and the literal content of Ben's use of the same sentence is *that waterboarding is not torture<sub>BA</sub>*. It follows that they are not affirming and denying the same proposition in their dispute.

The question we now want to raise is whether P&S's alternative account of the 'torture' example succeeds, where the shared meaning approach allegedly fails, in vindicating the thought that Anne and Ben are having a genuine disagreement. According to P&S, the disagreement is not over the propositions literally expressed by the speakers' sentences. Instead, the disagreement concerns which meaning or concept should play a particular *functional role* that they both associate with the term 'torture' (2013a, 21). Is it really plausible that Anne and Ben are genuinely disagreeing about which definition best fulfills a specified functional role?

In their discussion, P&S leave it vague just which functional roles Anne and Ben presumably associate with 'torture' in their policy debate:

Even if we suppose that the speakers mean different things by the word 'torture,' it is clear that we have not exhausted the normative and evaluative work to be done here. After all, in the context of discussions about the moral or legal issues surrounding the treatment of prisoners, there is a substantive question about which definition is *better*. (2013a, 19)

What exactly is this normative and evaluative work? What exactly are the moral and legal issues at stake? P&S don't say. But resolving that question is crucial to specifying what exactly Anne and Ben are disagreeing about—namely, which of the divergent definitions they respectively hold better fulfills the functional roles that they presumably associate in common with 'torture.'

Now, it is important to observe that principles of interpretive charity apply at the level of mental content as well as at the level of linguistic content. So, to paraphrase P&S, one crucial type of data in figuring out the content of an individual's mental states are facts about the individual's associated dispositions. It follows that if it were uncharitable to interpret Anne and Ben as picking out exactly the same meaning or concept in their uses of the term 'torture' given their divergent patterns of applying the term, then it would also be uncharitable to interpret them as picking out the same functional role given their divergent dispositions to see the specific moral and legal issues as implicated by their use of the term 'torture' in their current conversational context. To comply with P&S's *prima facie* interpretive constraint (applied to mental contents as well as to linguistic contents), we must respect Anne's and Ben's actual dispositions in attributing to them their

respective mental states that partly make up their disagreement. But which dispositions? One natural suggestion is that we should focus on Anne's and Ben's dispositions to identify precisely which legal and moral questions are at issue in their dispute: e.g., what implications does the ideal of human dignity have for classifying certain action types as torture; what is the relevance of the considerations of collective well-being or security in such classifications; what relevance, if any, does our enemies' treatment of our military personnel have for our conception of what constitutes torture; what is the relevance of the past practices of interrogation and the evolving collective standards of decency, both legal and other, in the world's civilized communities; and so on.<sup>8</sup> If there were a precise match in the two parties' dispositions in raising and answering such questions, then that match would provide good prima facie evidence that they had the same moral and legal functional roles in mind in their dispute about how to apply 'torture.' And this shared conception of the relevant legal and moral functional roles would then ground a substantive meta-level disagreement between Anne and Ben about which definition of 'torture' should play the relevant functional role.

However, it is highly unlikely that any two individuals will actually have precisely matching meta-level cognitive dispositions. A familiar feature of moral and legal arguments is that the parties to (what they take to be) an object-level dispute also diverge at the meta-level about how to demarcate the moral or legal roles at issue, about which epistemic considerations are most probative, or which practical issues hang in the balance. Prior to their discussion, Anne and Ben are likely to have divergent dispositions at the meta-level about the precise moral and legal issues at stake in characterizing torture. If we follow P&S's interpretive constraint and treat the individuals' dispositions as a prima facie guide to attributing thought contents, we should conclude that Anne and Ben do *not* have a meta-level disagreement about which definition should fill a specific moral and legal role.<sup>9</sup>

In short, if P&S's prima facie interpretive constraint undercuts positing any strict disagreement at the semantic level—in the contents literally expressed by speakers' words—then it does the same at the level of mental states and the functional roles that are presumably communicated pragmatically. For just as a precise match in the two parties' application-dispositions for 'torture' would be highly unlikely, a precise match in the functional roles they associate with 'torture' would be similarly unlikely. If we assume that rational incompatibility requires logical inconsistency, then P&S's claim that MLN can generate genuine disagreements in moral and legal cases seems unfounded.

Our argument in this section has been ad hominem. It has relied on an interpretive principle P&S appeal to but which we ourselves have not endorsed. We will in fact go on to discuss that principle's limitations in the next section. Our argument here is meant to highlight the incompleteness of P&S's proposal. In order to establish MLN as a possible, let alone plausible, explanation of genuine disagreements in the moral and legal domains, P&S need to explain much more carefully how MLN generates rationally incompatible thought contents.

At this stage, P&S might want to resist our reading of 'rational incompatibility' as requiring logical inconsistency. They might propose a weaker type of rational incompatibility and show that MLN can generate genuine disagreement in that weaker sense. We'll make two comments about this 'weakening' strategy. First, that strategy won't be easy. *Incompatibility* is a modal notion. It entails that two thought contents *cannot* be reconciled (in some sense). And *rational* incompatibility entails that the relevant incompatibility is not just accessible to some external

interpreter. It must be accessible from the perspective of the parties in the dispute. Logical inconsistency meets these demands. But it's not obvious how to define a weaker notion of rational incompatibility in a principled way. Second, once a weakened notion of rational incompatibility is in hand, P&S will need to explain the costs and benefits of switching from the more demanding notion of rational incompatibility prominent in metaethical debates to their preferred weaker one. For instance, giving up on the possibility of vindicating the logical validity of modus ponens would clearly be a high price to pay.

### 3. Coordinating and epistemic commitments

If P&S don't succeed in vindicating genuine substantive disagreement, then their challenge to the shared meaning explanation of moral and legal disagreements will be significantly diminished. Recall that their main goal in introducing MLN was to show that the shared meanings approach is not the only game in town for explaining how disputes involving normative terms could express substantive disagreements. If their account fails to vindicate genuine disagreement at the meta-level, then it does not provide a challenge to a shared meaning account of disagreement. Still, we agree with P&S that the shared meaning approach has not been sufficiently defended in the literature. In the remainder of the paper, we make a positive case for it. In this section, we highlight the coordinating epistemic commitments normally undertaken by participants in moral and legal debates using one of P&S's core examples. In the next section, we explain why these commitments support the ascription of shared meanings as the default interpretation of the participants' words.<sup>10</sup>

Let's return to P&S's 'torture' example. We can imagine the debate between Anne and Ben running as follows:

- (a) Anne: Waterboarding is torture.
- (b) Ben: No, waterboarding is not torture; it doesn't inflict pain rising to the level of death, organ failure, or the permanent impairment of a significant body function.
- (c) Anne: That's not what it takes to qualify as torture. What's relevant is that waterboarding inflicts severe suffering, physical or mental, in order to obtain information or to punish.

Can Anne and Ben associate the same meaning with the word 'torture' when their application dispositions for that word so diverge? As we have seen, P&S worry that ascribing shared literal contents in disputes like Anne and Ben's may violate what they take to be an uncontroversial prima facie constraint on interpretation, which we quote again:

Regardless of one's precise views in semantics, it should be uncontroversial that at least *one* crucial type of data for figuring out what a speaker means by a term T are facts about the speaker's usage of T—patterns of usage that reflect her disposition to apply that term one way or another, more generally. (2013a, 16)

We'll call this interpretive constraint 'RSAD,' short for 'Respect Speaker's Application Dispositions.' RSAD says that there is prima facie reason to take Anne's use of 'torture' to pick out an action kind that reflects her current dispositions for applying 'torture' and Ben's to pick out a kind that respects his alternative dispositions. In the absence of competing or additional



interpretive principles, we should conclude that they are literally expressing different contents with the term ‘torture,’ and that their claims pick out distinct action kinds. Moreover, in affirming or denying the sentence ‘Waterboarding is torture,’ Anne and Ben should be seen as each making *literally true* statements in their own respective idiolects since their statements accurately reflect their own respective idiosyncratic linguistic dispositions to apply the term ‘torture.’

We agree with P&S that RSAD is *one* important prima facie constraint on charitable interpretation. In the absence of competing considerations, it seems uncharitable to suppose an individual is systematically mistaken about the application conditions of her own words. But there are many additional constraints on charitable interpretation. We’ll start our search for these additional constraints by examining a strengthened version of RSAD, which we will call ‘RSAD\*.’ RSAD\* takes RSAD to be the *only* interpretive constraint.<sup>11</sup> As we’ll see, RSAD\* has strong counterintuitive consequences in the case of legal and moral terms. We’ll suggest that the reason RSAD\* is so counterintuitive is that it neglects certain coordinating and epistemic commitments that are central to our practice with moral and legal terms. These commitments, we’ll suggest, support as the default interpretation the ascription of the same contents to Anne and Ben’s use of ‘torture.’

A telling problem for RSAD\* is that it *makes truths about moral and legal questions much too easy*. Suppose a lawyer like Ben were to espouse an outlandish definition of ‘torture.’ He might declare that torture is simply the intentional killing of prisoners, and, consequently, that no mistreatment short of causing death would violate the legal prohibition of torture. Or, even more surprisingly, Ben might claim that torture is the act of dispensing candy to prisoners. These definitions are completely divorced from our moral and legal traditions of thinking about torture. But if we respect RSAD\*, each of Ben’s claims would count as true simply in virtue of his espousing the relevant definition. And the definition would determine the truth-conditions of any of Ben’s various moral and legal claims involving the term ‘torture.’

We take this consequence to be a *reductio* of RSAD\*. It’s not just that the definitions strike external observers as crazy; these definitions seem crazy even from Ben’s own point of view. In moral and legal debates, normal people do not take the definitions they espouse to determine exclusively the truth-conditions of their moral or legal claims. Normal moral agents recognize that their own application dispositions of moral and legal terms are *fallible*. Let’s assume Ben is currently disposed honestly to judge that waterboarding isn’t torture. Waterboarding doesn’t count as torture, he thinks, because it doesn’t seem so painful and prisoners don’t die from it. But even from Ben’s own point of view, this judgment could turn out to be mistaken.

Indeed, Ben may be disposed to revise his current first-order application dispositions in the light of new empirical information, armchair reflection, or interpersonal testimony and discussion. For instance, he might be disposed to change his mind about what constitutes torture when he learns about what it’s really like to undergo waterboarding and its lasting psychological impacts. Or he might be disposed to change his mind if he engaged in theoretical moral reflection on the dignity of persons or on agent-centered moral constraints against inflicting harm. Each of these changes is grounded in Ben’s own prior dispositions—dispositions to revise his first-order application dispositions. From Ben’s point of view, moreover, these higher-order dispositions are not dispositions to change the topic. Instead, these higher-order dispositions reflect his own prior epistemic commitments. They reflect his efforts to home in on the truth about the topic he was thinking and talking about all along. From Ben’s point of view, making these sorts of changes to

his first-order application dispositions would constitute epistemic progress. He would be *getting closer to the truth* about torture itself.<sup>12</sup>

Ben's fallibilist epistemic commitments may be obscured if we focus exclusively on specific conversational contexts. However, Ben's endorsement of a novel definition of torture is not just a one-off move in a conversational game with Anne in which he tries to get her to adopt his own application dispositions. Remember, we are assuming that Ben sincerely believes that waterboarding isn't torture. His definition reflects his own best judgment about what constitutes torture.<sup>13</sup> Normally, an individual's understanding of what's picked out by a term like 'torture' is tightly intertwined with that of others. Ben's formation of his concept was likely originally prompted by a public linguistic label (indicating a feature that is worth tracking), and his evolving understanding of that feature would be constantly susceptible to refinement through the use of that label in testimony and debate with others. Such epistemic integration with one's community is a core aspect of a normal individual's own understanding of specific objects, kinds, and properties. We presume that torture is an important but imperfectly understood feature of the world that we have all been talking about as 'torture.' So, Ben's endorsement of his new idiosyncratic definition presumably involves coordination of his current understanding of torture with that of his previous self and of others in his linguistic community. By his own lights, presumably, his new definition of torture was a refinement and correction of his previous understanding of *the topic that he'd been thinking about all along*. And Ben likewise would have taken this to be the topic that *others in his linguistic community* had been thinking about. If Ben is making an honest judgment in his discussion with Anne, he is not just trying to nudge or manipulate Anne's understanding of 'torture,' but instead trying to get at the truth about torture—i.e., the kind of action he and his community is picking out with 'torture.'

Let's examine more carefully at how the coordinating epistemic commitments we have highlighted—i.e., those aimed at epistemic coordination with one's previous self and with others on topics we imperfectly understand—structure the epistemology of the moral and legal domains. Consider the legal domain first. As a lawyer, Ben would be highly aberrant if he thought that his own personal dispositions alone established the correctness of his proposed legal definition of 'torture.' From his own lawyerly perspective, the constraints on a correct definition of the term include sensitivity to established laws concerning 'torture,' judicial precedents, the evolving standards of decency in civilized societies, and recognized interpretive strategies in law. If he is searching for a correct definition of 'torture,' Ben needs to make this definition acceptable within the legal tradition he's participating in, and not simply express his own private dispositions to classify cases. The ultimate validation of Ben's legal definition, even from his own perspective, depends on external constraints beyond his own application dispositions.<sup>14</sup>

We'd like to emphasize that our coordinating and epistemic commitments about definitions can be more or less demanding, depending on the topic in question and the conversational context. Consider P&S's example of a dispute over whether Secretariat, a racehorse, counts as an athlete (2013a, 16). Prima facie, determining precise boundaries for who or what exactly counts as an athlete and whether nonhuman animals are excluded is not something we care very much about. In our shared linguistic tradition, the term 'athlete' has normally been applied exclusively to humans who engage in sport. But for the purposes of sport, we don't really require stable interpersonal criteria for distinguishing athletes from nonathletes. Since nothing very substantial is at stake here, we can afford to sharpen our somewhat inchoate understanding of 'athlete' in different ways for different conversational purposes. At the Flemington Racecourse or Churchill Downs, we might be

happy to talk about racehorses as athletes, while resisting including them in another context where the comparison class is a list of famous human athletes. In short, when a domain of discourse does not require stable, relatively determinate intensions, it's natural to treat the terms as having context-variable intensions that reflect the interests and circumstances of specific interlocutors.

The situation is different for legal terms. Determining precisely which actions count as 'torture' within a specific legal jurisdiction is of considerable public interest to governments, judges, lawyers, soldiers, prison administrators, and victims of abuse. And for legal purposes, the precise extension of 'torture' can't be sharpened differently in different conversational context to suit the specific interests of interlocutors. A central role of a system of laws is to provide *(relatively) determinate action-guiding norms that remain stable across conversational contexts*. In general, if a lawyer honestly believes that his definition of a legal term is true, he is committed to its truth holding for the whole legal jurisdiction—and in the case of terms like 'torture,' even across jurisdictions—and not just for himself.<sup>15</sup> As a lawyer, therefore, Ben needs to argue that his proposed definition (imposing suffering at the level of organ failure or death) is the correct standard for classifying *all* cases of torture in *all* contexts relative to the designated legal jurisdiction, and even across jurisdictions. And he will need to do so by appealing to elements of the legal tradition he is part of—established laws, judicial precedents, prevailing practices in civilized societies, certain moral principles, and so on. This legal argument would not be just a way of convincing others; it would also be how Ben convinces himself that his legal definition is correct.

In fact, it's natural to construe the legal system within a jurisdiction as engaging in a distributed but coordinated epistemic project. The system as a whole seems to work to identify a *determinate extension that is mutually justifiable on the basis of the shared legal tradition*. Individuals within the legal system can be seen as epistemic agents seeking to identify determinate stable verdicts about cases that will survive the process of reflective critique.<sup>16</sup> Like the scientific system of experimentation and peer review, the legal system can be characterized as a distributed epistemic system designed to foster diversity and competition in finding the most compelling, mutually justifiable verdicts. And like the scientific system, the legal system is fallible in achieving that result. In both cases, poorly justified verdicts or mistakes may be endorsed and even entrenched within the system; and, over time, they may generate new sets of standards for mutual justification that are unjustifiable but for the initial mistakes. Perhaps legal systems lack the kind of self-correcting mechanisms or capacities that scientific systems have. Or perhaps their self-correcting mechanisms or capacities are comparable to those of scientific systems.<sup>17</sup> This is an issue that would require a deep dive into the nature of legal systems and into the nature of science as well, and we opt to bracket it here. Either way, a legal system seems to play a distinctive epistemic role—namely, to identify, systematize, and maintain a set of determinate, mutually justifiable legal verdicts.<sup>18</sup>

We won't look at the moral case in analogous detail here, but we think it has important similarities with the legal case. One major difference, of course, is that moral claims are not relativized to local jurisdictions. If Ben is wondering whether waterboarding is *morally permissible*, he is looking for a verdict whose validity is not restricted in its scope to a locality of any sort. Another difference is that moral inquiry is much less institutionalized than legal or scientific inquiry. There is no well-demarkated group of moral experts whose deliberations and verdicts are generally recognized as authoritative. There is also no set of institutions that have the capacity to change the moral status of at least some action types, or that are in charge of issuing

verdicts that are stipulative in the sense of drawing lines when the applicable considerations do not warrant decisive verdicts. Still, we think there are key similarities in the structure of moral and legal inquiry and debate. As in the case of a legal system, a central role of a morality, or a set of moral norms, is to provide (*relatively*) *determinate, action-guiding norms that remain stable across conversational contexts*. Just as in the case of legal verdicts, if Ben claims that this particular lie is morally wrong, he is committed to holding that relevantly similar speech acts are also morally wrong, and to taking such moral verdicts to be stable across time, points of view, and conversational contexts. Thus, on the key issue of speakers' commitment to the stability and determinacy of extension, 'morally wrong' seems much closer to legal terms than to 'athlete' in P&S's example.

## 4. Shared meanings as the default

We are now in a position to answer P&S's challenge. Why posit *shared meanings* to explain moral and legal disputes? Our answer, in a nutshell, is that individual speakers' own coordinating epistemic commitments are what support treating shared meanings as the default semantic interpretation for moral and legal terms. And a charitable interpretation of individuals' words should seek to vindicate this aspect of their epistemic perspective.<sup>19</sup>

In the preceding section, we used RSAD\* as a foil for articulating different aspects of our practice with moral and legal terms. From the perspective of individual speakers, one's current applications dispositions and explicit definitions are *fallible* about the true instantiation conditions of moral or legal kinds.<sup>20</sup> Moreover, from a normal speaker's perspective, one's own ignorance and error can be corrected through suitable epistemic processes. In the moral and legal domains, the source of such correction is not just armchair reasoning; corrections are often motivated by testimony, debate, and interpersonal calibration with those engaged in shared moral and legal practices. In short, normal people think they can get closer to the truth about the instantiation conditions of the moral and legal properties their own words pick out in part by reasoning with others. This commitment to a coordinated epistemic practice makes sense whenever we take our words to pick out stable features of the world that are relevant to our community. Two heads are usually better than one. But epistemic coordination is particularly compelling in the moral and legal domains, where we are seeking to identify action-guiding norms of conduct for everyone—all agents or all agents in a jurisdiction. For these purposes, it's important that our moral and legal terms pick out (*relatively*) determinate extensions that remain stable across different conversational contexts and that can be publicly justified on the basis of the community's past moral or legal practices.<sup>21</sup>

The upshot, then, is that individual speakers' own epistemic dispositions and commitments seem to presuppose that moral and legal terms have stable extensions over time and between individuals. These epistemic commitments to stability are treated by speakers themselves as an important constraint in correcting their fallible patterns of understanding. A charitable interpretation, therefore, should seek to assign semantic contents that respect individuals' own commitment to shared meanings.

The reason to favor shared meanings, on this account, is not just a bare intuition that speakers genuinely disagree in moral and legal disputes, or a 'feeling of conflict' between the speakers (P&S 2013a, 25), or an intuition that such disputes matter in ways that merely verbal disputes generally do not. These data points fail to capture how deeply dispositions for interpersonal

coordination and epistemic self-correction are embedded within individuals' own ongoing linguistic practices with moral and legal terms. These epistemic dispositions are not merely a set of data points; our moral and legal linguistic practices would be unrecognizable if they were not epistemically integrated with social practices the ways we've highlighted. Arguably, our epistemic dispositions to treat moral and legal questions as targets of mutual interest and shared inquiry is what grounds the isolated intuitions about substantive disagreement highlighted by P&S.

It's worth stressing that this epistemic commitment to shared meanings is not best construed as simple deference to conventional patterns of understanding entrenched within a community. According to our epistemic norms, entrenched ways of understanding what's morally permitted, for example, are corrigible in the light of further reflection or new empirical information (e.g., about the psychology of patients of the relevant action kinds). Intuitively, this correction gets us closer to the truth about what we were interested in all along—moral permissibility. *Mutatis mutandis* for our epistemic norms governing currently accepted legal doctrine. The epistemic practice of lawyers and judges does not require them to accept simply whatever most legal philosophers currently believe, or to defer simply to currently entrenched judicial opinion. Good legal theorizing can attempt to vindicate or correct currently accepted legal interpretation on the basis of arguments and evidence. Thus, the default presumption of sameness of meaning is not a presumption in favor of deferring to current prevailing ways of understanding a term, which would amount to a commitment to moral or legal conventionalism.

In effect, we've argued that there is an important interpretive constraint that requires charitable interpretation to respect individuals' *epistemic dispositions* when assigning literal contents to their words. Let's call this constraint 'RSED,' short for 'Respect Speaker's *Epistemic Dispositions*.' The force of our argument so far, put simply, is that other things being equal, RSED should trump RSAD—i.e., the commitment to respect individuals' current *application dispositions*. And insofar as individual speakers' epistemic dispositions for moral and legal terms presuppose interpersonal coordination, RSED supports a strong default interpretive presumption in favor of positing shared meanings.<sup>22</sup>

Of course, this default presumption in favor of shared meanings can be defeated if individuals' epistemic commitment to their own substantive understanding of a term is more important to their overall practice than their commitment to epistemic coordination with their community. But our aim here is not to establish that our actual use of moral and legal terms is *always* univocal. Our aim is more modest. We aim to show why the shared-meaning view is to be preferred, *other things being equal*, in the case of moral and legal terms.<sup>23</sup> We would need strong independent reasons to override the subjects' own coordinating epistemic dispositions. A simple divergence in speakers' current application dispositions for a word—as in the dispute between Anne and Ben—will not ordinarily suffice to defeat the default interpretive presumption in favor of positing a shared meaning because it does not suffice to undermine their commitment to shared epistemic standards.

Now, a defender of MLN for moral and legal terms might point to the difficulty of finding a univocal, mutually justifiable interpretation. P&S call this 'The Shared Meaning Task,' and they argue that it is a point in favor of MLN that it cleanly avoids the need to find an interpretation that makes sense of the divergent commitments of the members of a linguistic community (2013a, 25). We agree that finding a satisfying univocal interpretation is often difficult. And if it should prove impossible to justify a single extension for moral and legal terms within a community, this *might* be a reason to favor supposing that individuals express distinct idiolect meanings with their use of

a term. However, there are reasons to be cautious about this conclusion. First, any difficulty of finding a univocal interpretation does not just apply to public language, it also applies at the level of idiolect meaning and thought content, as we argued in section 2 above. Individuals' understanding of moral and legal terms is complex, with partial and competing application dispositions, inferential dispositions, deferential and coordinating dispositions, epistemic and methodological dispositions, and so on. Given that complexity, the task of specifying the idiolect meaning individuals associate with contested terms like 'torture' may be just as difficult as the Shared Meaning Task. Second, there are many different types of semantic contents one might appeal to in the Shared Meaning Task. If it proves impossible to justify a univocal reference for 'morally right' or 'torture,' charitable interpretation might favor univocal fictionalism, contextualism, relativism, or expressivism. Unlike MLN, these univocal interpretations vindicate individual speakers' commitments to stable diachronic and interpersonal coordination with their use of moral and legal terms.

## 5. Concluding remarks

We'll conclude with one observation about P&S's proposal to construe disagreements as MLNs. We suspect that the coordinating epistemic commitments we have highlighted in the preceding two sections are typically strong enough that even disputes that start out as MLNs will tend to turn into disputes with shared literal contents. When trying to rationally adjudicate a serious issue, we will normally seek to formulate the dispute in such a way that the real point at issue is literally conveyed by our words. Speakers would need strong and special independent reasons to continue a dispute as an MLN even after learning that they associate different meanings with the key terms used in the dispute, and after recognizing that the substantive issues at stake (about which definition is best for certain purposes) are not directly expressed by the words they have been using in the MLN. Any such independent reasons, we further suspect, would tend to compromise the integrity of those disputes.<sup>24</sup>

Consider for example the recent dispute among US senators about the definition of 'infrastructure' (Smith 2021). The real bone of contention between Democrats and Republicans was whether the Senate should pass certain education, health, eldercare, and other measures. Knowing that full well, both sides opted to fight over the definition of 'infrastructure' rather than debate the pros and cons of the relevant laws directly. In so doing, Democrats were trying to manufacture a sense of urgency for the education, health, eldercare, and other measures they favor, which they called 'soft infrastructure,' by bundling them together with the measures for 'physical infrastructure,' such as roads, bridges, ports, and broadband. And Republicans wanted to avoid arguing directly against the former measures that are generally popular with the electorate. Thus, the maintenance of a patently disingenuous and frustrating dispute about what 'infrastructure' means. If the participants were more honest with each other and the public, the dispute would have quickly turned into a dispute couched in terms of shared literal meanings about the virtues and vices of the relevant measures and would have been less inclined to manipulate and exploit the prevailing psychological associative patterns.

Similar observations could be made about most, if not all, of P&S's examples. For our parts, we have a hard time imagining what could have motivated two speakers to persist in an MLN about the right meaning of 'athlete,' and whether Secretariat falls within the extension, unless the speakers' primary goal was to fill up an allotted radio time with inconsequential and diverting

gibberish.<sup>25</sup> We would have similar difficulty fathoming the motivations of two speakers who persist in debating about what ‘chips’ really mean—French fries or crisps (P&S 2013a, 27)—instead of switching quickly to talk, for example, about how the menu should be written to minimize customers’ confusion or disappointment. Any such persistence would be material for a sitcom scene, and a lame one at that, but not for real life.

P&S (2013b) have also argued that we should characterize many fundamental legal disputes—e.g., between appellate judges—as MLNs about the meanings of key legal terms. Anticipating the objection that such a characterization would do a poor job of distinguishing between disputes about what the law *is* from disputes about what the law *should be*, P&S (2021, 163–64) argue that we should not be squeamish about blurring that distinction. It is the erroneous folk belief that judges should never make new law, and should only apply pre-existing law, they argue, that motivates judges in many cases to behave as if their disputes are about what the law is, rather than about what the law should be. We ourselves believe that the distinction is much more important and robust than P&S think. And while cognizant that judges, like everyone else, sometimes act disingenuously (or confusedly) in conducting and characterizing the disputes with each other, we suspect that the instances are much rarer than what P&S’s diagnosis implies.<sup>26</sup> However, setting these differences aside, what is important to highlight here is that even in this set of P&S’s examples, something has to have gone awry for speakers to persist in their MLNs when they are fully apprised of the metalinguistic nature of their disputes and their real first-order stakes. When speakers are afforded abilities to be upfront and their primary goal is to find the right answers, even those disputes that start out as MLNs, we surmise, will quickly turn into disputes conducted via literal semantic contents.

P&S suggest that theoretical parsimony favors explanations of disagreement in terms of MLN. Some linguistic disputes clearly do involve MLN. So any total theory of communication must have the theoretical resources to explain such disputes. But explaining linguistic disputes in terms of shared meanings requires more theoretical resources (2013b, 268). So why should we posit shared meanings over and above MLN? We acknowledge that MLNs do take place. But our view is that they are very different from moral and legal disputes that involve epistemic coordination of the sort we have delineated. The price for theoretical parsimony, we suggest, may be that large parts of common-sense epistemology must be reconstrued as purely pragmatic clashes. Another way of putting this point is to observe that serious disputes that start out as metalinguistic disputes tend to be converted, where possible, into object-level disputes conducted through literal semantic contents, for it is in the context of such disputes that we avail ourselves of the time-tested and hard-won epistemological resources that we have come to equip ourselves.

---

## Acknowledgments

We’d like to thank David Plunkett and Tim Sundell for extensive discussion and feedback. Thanks also to Nick Allott, Eliot Michaelson, Bill Watson, and several anonymous reviewers.

François Schroeter is associate professor in the philosophy discipline at the University of Melbourne. His work focuses on metaethics, normative concepts, and moral psychology.

Laura Schroeter is associate professor in the philosophy discipline at the University of Melbourne. Her work focuses primarily on concepts, reference, two-dimensional semantics, and metaethics.

Kevin Toh is a professor of philosophy of law at the Faculty of Laws, University College London. His research focuses on topics in philosophy of law, philosophical ethics, and constitutional theory.

## Notes

<sup>1</sup> By ‘legal philosophy’ and like terms, we shall in this paper be referring to the subfield of legal philosophy that deals with the nature of law, which sometimes goes by the name ‘general jurisprudence.’

<sup>2</sup> Thus, an MLN is a negotiation over *which meaning a term should have* in a conversational context. For ease of exposition, we shall sometimes use ‘MLN’ to refer also to theories that explain what is at stake in normative disagreements in terms of MLN.

<sup>3</sup> P&S also make more positive claims in favor of MLN: e.g., “[W]e aim to demonstrate that the metalinguistic analysis is a strong, if not inevitable, candidate in the explanation and analysis of *any* particular domain of normative or evaluative discourse” (2013a, 4); and MLN is “particularly plausible as an analysis of many normative and evaluative disputes” (2013a, 7). P&S’s positive case for MLN seems to rest largely on the theoretical parsimony of a purely pragmatic account of disagreement. We will not address this issue here, except for a few remarks at the end of section 5.

<sup>4</sup> See (P&S 2013a, 11). We’ll return to the question of rational incompatibility in the next section.

<sup>5</sup> As P&S note, this transition is not explicitly argued for in the metaethics and legal philosophy literature; it is simply presupposed. P&S suggest theorists might appeal to inference to the best explanation or some other types of argument to shore up that claim (2013b, 253).

<sup>6</sup> For similar worries about whether MLN can vindicate genuine disagreements, see (Cappelen 2018, 175–77).

<sup>7</sup> See the helpful summary of such attempts by metaethicists in van Roojen (2018, sec. 4.1). A related discussion of the problem as it applies to legal expressivism can be found in (Toh 2005, sec.10).

<sup>8</sup> We’ve phrased these questions in a way that presupposes that the two speakers share the concepts involved in formulating them. But notice that the questions themselves involve abstract contents. And on P&S’s approach, it’s an open question whether the two parties associate the same concepts with terms like ‘human dignity’ or ‘well-being.’ This is an additional indication of the sheer improbability (about to be discussed in the text) of there being a precise match in the two parties’ thought contents about of the functional roles that are at issue in defining ‘torture.’

<sup>9</sup> One might be tempted to appeal to the speakers’ shared context of utterance, or their broader social context, and argue that the two parties to the dispute should be interpreted as strictly disagreeing on the basis of *whatever functional roles are determined by their shared context*. Two points are relevant here. First, if speakers’ *matching conceptions* of their shared context (rather than the context itself) is meant to secure that the speakers pick out precisely the same role, then the problem of getting a precise match will just crop up in a different place—i.e., their dispositions to identify which aspect of their context is relevant to fixing the functional role at stake. Second, if the external context itself helps to secure a precise match in speakers’ thoughts about the functional role played by ‘torture,’ it’s not clear why it can’t play an analogous role in securing a precise match in the object-level contents the two speakers express with ‘torture.’ If anything, facts about speakers’ sociolinguistic context seem more directly relevant to interpreting elements of their *linguistically formulated claims* (‘waterboarding is torture’) than to interpreting the intentional content of *implicit dispositions* to treat specific functional roles as probative (‘such-and-such moral/legal functions are at issue in defining the word ‘torture’).

<sup>10</sup> Of course, it’s possible to reject the claim that moral and legal disputes involve *any* genuine disagreement at any level. But this interpretation needs a positive defense just as an interpretation that posits shared meanings does. In the following two sections, we argue that individuals’ epistemic practices with moral and legal terms provide strong *prima facie* support for interpreting



members of the same communities as using those terms with a shared meaning. To interpret moral and legal debates as involving individuals systematically talking past each other and focusing on their own private epistemic standards does not fit with our reflective epistemic commitments in the moral and legal domains. However, we'd like to emphasize that our argument is based on *the epistemology of these specific domains*. Some theorists argue that *fundamental metaphysical disputes* may involve systematically talking past each other (e.g., Hirsch 2005; Thomasson 2017; Belleri 2018). We take no stand on this issue here since debates about, say, 'object' or 'exists,' may not involve a strong commitment to epistemic coordination with one's community.

<sup>11</sup> RSAD\* is close in spirit to MUMPS, Tim Sundell's (2012) approach to semantic interpretation. However, we are not assuming here that P&S's defense of MLN is committed to MUMPS.

<sup>12</sup> Of course, not every higher-order disposition is an epistemic disposition. Ben might also have higher-order dispositions to change his way of applying 'torture' if paid large sums of money by a lobbying firm. Or he might be motivated by peer pressure or prospects of promotion to change his application dispositions. Our point is simply that there are some higher-order dispositions that reflect good epistemic reasons, and that this is crucial to the thinker's own point of view on the reference and satisfaction-conditions of their words.

<sup>13</sup> See (P&S 2021, 19) for a similar assumption.

<sup>14</sup> Ben also should not fashion a definition merely to make his client's behaviour count as legally permissible. These sorts of considerations, of course, might be part of what actually *motivates* him to propose a definition. As P&S conjecture at one point:

During much of the Bush administration, it was more or less agreed that the sentence "the United States doesn't torture" should come out true. So a lot rested on how exactly 'torture' was defined, and the Bush Justice Department bent over backwards to find some definition, no matter how strained, that excluded waterboarding. (2021, 20).

But by Ben's own lights, that type of consideration is not what makes his definition *true*. Though such a consideration could count as an external constraint in a sense, it would be an external constraint of a wrong sort. It is a *purely pragmatic* reason to accept a definition, not an epistemic one.

<sup>15</sup> Of course, many legal definitions are not stable across jurisdictions. What counts as defamation in England, for example, may not be defamation in the US. Analogous phenomena can be observed sometimes for notions that figure in different areas of the law of a single jurisdiction. For example, 'consideration' has slightly different meanings in the English laws of contracts and restitution. And 'dishonesty' had slightly different meanings in criminal and civil legal contexts until the UK Supreme Court opted for a uniform definition in *Ivey v. Genting Casinos (UK) Ltd*, [2017] UKSC 67. Though *Ivey* supports our line of thinking, the facts are complicated enough to warrant a limited hedging of the sort we do in the text, most explicitly in the penultimate paragraph of section 4 below. Thanks to Niamh Connolly and Mark Dsouza for instructive discussions of these matters.

<sup>16</sup> Even judges, whose verdicts have a provisional status of setting standards and precedents, are usually required to provide justifications, and their verdicts are subject to critique and rejection by some combination of other judges (sitting in higher courts or in courts deciding later decisions), commentators, the public, and legislators. Therefore, even judges are motivated to find justifiable verdicts that will withstand scrutiny.

<sup>17</sup> For a recent Whiggish characterization of legal systems that could go some way to sustain a latter assessment, see Railton (2019).

<sup>18</sup> One might wonder why a proponent of MLN couldn't appeal to this idea that individual speakers are committed to taking their disagreement about how to define legal terms as a disagreement about which definition best meets the standards of mutually justifiable legal interpretation. Our answer is twofold. First, our foregoing ad hominem argument against P&S applies here. There is no reason to think that different individuals have exactly the same substantive understanding of the standards governing legal interpretation. Instead, individuals are committed to *finding common ground* with others in their actual legal tradition. Thus, P&S's individualistic principle of semantic interpretation on the basis of matching dispositions (RSAD) will not vindicate genuine disagreement at the meta-level in this case. Second, and more importantly, by the individual's own lights, the standards of legal interpretation are *epistemic standards* for identifying legal truths. So interpretation should construe the parties to the dispute as coordinating on the same contents *at the object level* about legal facts rather than at the meta-level about appropriate standards of legal interpretation. Thanks to an anonymous referee for pushing us to clarify this point.

<sup>19</sup> Notice that in emphasizing the role of an individual's *epistemic commitments for self-correction*, our account of charitable interpretation departs significantly from a traditional Quinean approach which asks us to maximize truths or a Davidsonian approach which seeks to rationalize an agent's immediate actions (Quine 1960; Davidson 1984). These approaches favor interpretations that seek to vindicate the individual's *current application dispositions*. Our approach to charity, in contrast, seeks to vindicate the individual's own *reflective epistemic commitments*. We claim that this approach better reflects the individual's own perspective on the meaning and reference of their words. See Schroeter and Schroeter (2015) for further discussion of that point. Some proponents of different meanings in metaphysics have appealed to the Davidsonian account of charity to justify their interpretations (e.g., Hirsch 2005; Thomasson 2017). Again, we don't take a stand on metaphysical disputes, but we believe that the Davidsonian approach distorts the interpretation of moral and legal terms.

<sup>20</sup> In this paragraph and for much of the balance of this section, we will be assuming a regular invariant realist semantics of moral and legal terms. But as we note at the very end of the section, analogues of the considerations we delineate for positing shared meanings apply for alternative semantic views.

<sup>21</sup> This justification for attributing shared meanings is similar in spirit to Tyler Burge's defence of anti-individualism about concepts (Burge 1979, 1986, 1989). For details about our preferred approach to moral and legal meanings and concepts, see Schroeter and Schroeter (2014, 2015) and Schroeter, Schroeter, and Toh (2020).

<sup>22</sup> The observations we have marshalled in support of the shared meaning approach are not novel. Semantic externalists like Putnam (1975), Burge (1979), Kripke (1980), and Millikan (1984) have long stressed that individuals can be ignorant or mistaken about the application conditions of their own words and concepts. Our point is that this 'externalist' commitment to fallibility is part of the epistemic perspective of individual speakers themselves on the contents of their own words and thoughts. The motivation for semantic externalism is grounded in the first-person epistemic perspective and practices of individual thinkers themselves. This point is worth emphasizing because externalism is sometimes thought to be motivated by the semantic intuitions of external interpreters. To be sure, the famous thought experiments of Putnam, Kripke, and Burge test our interpretive intuitions about possible cases, but when we judge that Oscar is referring to H<sub>2</sub>O or that Bert is thinking about arthritis, despite their misunderstandings of those kinds, we are implicitly endorsing (what we take to be) the epistemic perspective of the individuals involved. We assume that Oscar is prepared to learn about the underlying nature of the stuff he calls 'water,' and

that Bert is prepared to stand corrected about what counts as ‘arthritis.’ This empirical corrigibility is a natural part of our common-sense epistemic practices. These practices, *not* any abstract theoretical commitment to causal theories of reference or reference magnetism, are what support shared meanings as the default interpretation.

<sup>23</sup> To be clear: our suggestion that the shared meaning view should be preferred other things being equal, or taken as the default interpretation, is consistent with the claim that *some* moral and legal disputes may be best interpreted as MLNs.

<sup>24</sup> Our discussion in this section concerns the *typical dynamics* of disputes that may start as MLNs. We have no commitment here to the claims that all MLNs are trivial or not worth having. For rebuttals of these last claims in the case of fundamental metaphysical disputes, see, e.g., Balcerak Jackson (2014), Belleri (2017), and Thomasson (2017).

<sup>25</sup> See the instructive discussion of ‘bull session’ in Frankfurt (1986).

<sup>26</sup> P&S (2021, 164n8) cite H. L. A. Hart in support of their claim that the folk belief is wrong and that judges often make new law. Dworkin (1967, 1972, 1975) disputed Hart’s ([1961] 1994, ch. 7) views about the extent to which judges exercise discretion, and sought to devise a theory of law that vindicates the phenomenology that judges have that even in many ‘hard cases’ their decisions are fully constrained by what the law is. In effect, we are siding with Dworkin on adjudicative phenomenology, and discounting the possibility pushed by P&S (2013b), according to which judges frequently view their disputes in hard cases as metalinguistic negotiations that imply what the law should be, and invariably pretend, in such cases, that they are disagreeing about what the law is.

## References

- Balcerak Jackson, Brendan 2014. “Verbal Disputes and Substantiveness.” *Erkenntnis* 79: 31–54.
- Belleri, Delia. 2017. “Verbalism and Metalinguistic Negotiation in Ontological Disputes.” *Philosophical Studies* 174: 2211–26.
- Belleri, Delia. 2018. “Two Species of Merely Verbal Disputes.” *Metaphilosophy* 49: 691–710.
- Burge, Tyler. 1979. “Individualism and the Mental.” *Midwest Studies in Philosophy* 4: 73–121.
- Burge, Tyler. 1986. “Intellectual Norms and Foundations of Mind.” *Journal of Philosophy* 83: 697–720.
- Burge, Tyler. 1989. “Wherein Is Language Social? In *Reflections on Chomsky*, edited by Alexander L. George, 175–91. Oxford: Blackwell.
- Burgess, Alexis, Herman Cappelen, and David Plunkett 2020. *Conceptual Engineering and Conceptual Ethics*. Oxford: Oxford University Press.
- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Dworkin, Ronald M. 1967. “The Model of Rules I.” In *Taking Rights Seriously* (1977). Cambridge, MA: Harvard University Press.
- Dworkin, Ronald M. 1972. “The Model of Rules II.” In *Taking Rights Seriously* (1977). Cambridge, MA: Harvard University Press.
- Dworkin, Ronald M. 1975. “Hard Cases.” In *Taking Rights Seriously* (1977). Cambridge, MA: Harvard University Press.
- Frankfurt, Harry. 1986. “On Bullshit.” *Raritan Quarterly Review* 6: 81–100.
- Hart, H. L. A. (1961) 1942. *The Concept of Law*. 2nd ed. Oxford: Oxford University Press.

- Hirsch, Eli. 2005. "Physical-Object Ontology, Verbal Disputes, and Common Sense." *Philosophy and Phenomenological Research* 70: 67–97.
- Kouri Kissel, Teresa. 2021. "Metalinguistic Negotiation and Logical Pluralism." *Synthese* 198: 4801–12.
- Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Basil Blackwell.
- Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Plunkett, David, and Tim Sundell. 2013a. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13 (23): 1–37.
- Plunkett, David, and Tim Sundell. 2013b. "Dworkin's Interpretivism and the Pragmatics of Legal Disputes." *Legal Theory* 19: 242–81.
- Plunkett, David, and Tim Sundell. 2021. "Metalinguistic Negotiation and Speaker Error." *Inquiry* 64: 142–67.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning.'" *Minnesota Studies in the Philosophy of Science* 7: 131–93.
- Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge MA: MIT Press.
- Railton, Peter. 2019. "We'll See You in Court!" In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 1–22. Oxford: Oxford University Press.
- Schroeter, François, Laura Schroeter, and K. Toh. 2020. "A New Interpretivist Metasemantics for Fundamental Legal Disagreements." *Legal Theory* 26: 62–99.
- Schroeter, Laura, and François Schroeter. 2014. "Normative Concepts: A Connectedness Model." *Philosophers' Imprint* 14 (25): 1–26.
- Schroeter, Laura, and François Schroeter. 2015. "Rationalizing Self-Interpretation." In *Palgrave Handbook of Philosophical Methods*, edited by Chris Daly, 419–47. London: Palgrave Macmillan.
- Smith, Noah. 2021. "Defining 'Infrastructure' Is a Nonsense Argument." *Bloomberg*, April 9. <https://www.bloomberg.com/opinion/articles/2021-04-09/the-meaning-of-infrastructure-is-a-pointless-debate#xj4y7vzkg>.
- Sterken, Rachel Katharine. 2020. "Linguistic Intervention and Transformative Communicative Disruptions." In *Conceptual Engineering and Conceptual Ethics*, edited by A. Burgess, H. Cappelen, and D. Plunkett, 417–34. Oxford: Oxford University Press.
- Sundell, Tim. 2012. "Disagreement, Error, and an Alternative to Reference Magnetism." *Australasian Journal of Philosophy* 90: 743–59.
- Thomasson, Amie. 2017. "Metaphysical Disputes and Metalinguistic Negotiation." *Analytic Philosophy* 58: 1–28.
- Toh, Kevin. 2005. "Hart's Expressivism and His Benthamite Project." *Legal Theory* 11: 75–123.
- van Roojen, Mark. 2018. "Moral Cognitivism vs. Non-Cognitivism." In *The Stanford Encyclopedia of Philosophy* (Fall), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/fall2018/entries/moral-cognitivism/>.