

The theory theory of metalinguistic disputes

Erich Rast, New University of Lisbon

Correspondence Nova Institute of Philosophy, Sala 119, Campus Campolide, Faculty of Social and Human Sciences, New University of Lisbon, Avenida Berna 26 C, 1069-061 Lisbon, Portugal
Email: erast@fesh.unl.pt

Funding information

Work on this article was conducted under the postdoctoral fellowship DL 57/2016/CP1453/CT0002 funded by the Portuguese Foundation for Science and Technology (FCT).

Abstract

According to the theory theory of metalinguistic disputes, disagreements in metalinguistic disputes are based on diverging underlying theories, opinions, or world views. An adequate description of metalinguistic disagreement needs to consider the compatibility and topics of such theories. Although topic continuity can be spelled out in terms of measurement operations, it is argued that even metalinguistic disputes about a term used in different, mutually compatible theories can be substantive because the dispute is indirectly about the virtues of the underlying theories. The account is defended against externalist and holist objections.

KEYWORDS

metalinguistic negotiations, verbal disputes, theory-contextualism, adequacy of definition, theoretical virtues, theory comparison

1 INTRODUCTION

The central thesis of this article is that metalinguistic disputes are almost always substantive because they are based on disagreements about the adequacy of underlying theories. Different theories indirectly characterize the meaning of predicates and general terms in different ways. Therefore, it is principally hard to distinguish substantive from merely verbal disputes in a post-Quinean setting that does not allow for a clear-cut analytic/synthetic distinction. I suggest that the *theory theory of metalinguistic disputes* correctly explains why many of these disputes are substantive, lay out criteria for topic continuity, and defend the approach against strongly externalist objections by authors such as Cappelen and Lepore (2005) and Cappelen (2018). In the end, substantive disputes are not merely about the adequacy of terms or lexical effects, but rather about the virtues and merits of theories and world views. According to the theory theory, a metalinguistic dispute about a term α is substantive if and only if the underlying theories are important for a group of people as a whole, regardless of what the individual speakers think is important about α . What Plunkett and Sundell (2013) call the social role of a term α might be, but does not have to be among the contributing factors that make an α -theory important.

The remainder of this article is structured as follows. In Section 2, the approach is laid out and motivated. In Section 3, several objections are defused, and a summary is given in Section 4.

* This is a final draft version with a few corrections missing. Pagination differs from the published original. For quotations, please refer to the full article published as Erich Rast (2020): The Theory Theory of Metalinguistic Disputes, publ. online (early view), *Mind & Language*, <https://doi.org/10.1111/mila.12355>, 1-19.

2 THE THEORY THEORY

From now on, I will refer to the proposed approach as the *theory theory*. It is intended to explain synchronic metalinguistic disputes between interlocutors, as in Burgess and Plunkett (2013ab), Plunkett and Sundell (2013), and Plunkett (2015). It is also supposed to explain the diachronic perspective on meaning change (or lack thereof), as it is discussed by Carnap (1950), Strawson (1963), Cappelen (2018), Sawyer (2020), and in the literature on semantic amelioration such as Haslanger (2012).

In a synchronic scenario, the theory theory states that when two agents disagree about terms, they do so based on two different theories or world views A and B. In a diachronic scenario, the meaning of a term changes or is changed, and according to the theory theory this goes hand in hand with a change from an old theory A to a new theory B.

2.1 The loose sense of “theory”

In what follows, the term “theory” encompasses all kinds of related entities like world views, sets of beliefs, and parts of common-sense ontologies. These are theories in a loose sense of the word, and they are not required to be *scientific* theories. This is a liberal take on the word “theory”, encompassing various related conceptions, but there are some commonalities between them. First, theories can be described as sets of sentences or formulas in a logical language, and a theory change can be considered a transition from one set of formulas to another. Second, theories can be local, which means that a theory may be part of a larger theory, sets of beliefs, or world view, and a theory change need not necessarily affect other theories and beliefs. For example, if someone’s theory of atoms changes, then this need not affect that person’s “theories” or opinions about Paris. Generally speaking, a theory change may or may not affect other theories, world views, opinions, and beliefs. This shall be called the *locality thesis* from now on. Section 3.3 deals with an argument against it. Until then, the locality of meaning changes will be assumed without further justification. Third, we can keep track of theories, whether our own or those proposed by others, without endorsing them, and we can investigate two theories A and B and their consequences without running into paradoxes even when the union of A and B is contradictory. For example, this ability enables someone to understand this paper and consider its philosophical merits without endorsing all its claims and even while explicitly disagreeing with some of them. We are contrasting and comparing competing theories all the time, and while we do this we do not necessarily need to changes other theories, beliefs, and opinions we contemplate or endorse.

2.2 Characterizing metalinguistic disputes

What are metalinguistic disputes? As a rule of thumb, these are disputes about the meaning of terms or about how to adequately use a term in a context.

Plunkett (2015) distinguishes further between descriptive and normative metalinguistic disputes. A descriptive metalinguistic dispute concerns what a word means in public language, whereas a normative metalinguistic dispute concerns what a word should mean. Plunkett (2015, pp. 833-834) gives an example from Barker (2002) to illustrate that often both readings are available when speakers disagree with each other. A dispute about an utterance of “Feynman is tall” can be

based on the question of whether Feynman fulfills the contextually provided standards for being tall, which is a descriptive issue, or it may be based on a disagreement about what counts as a relevant standard for being tall.

Another important distinction is whether such a dispute is explicitly or only indirectly about a linguistic expression. While Burgess and Plunkett (2013ab) are primarily concerned with disputes in which speakers explicitly disagree about the meaning of linguistic expressions, Plunkett and Sundell (2013) and Plunkett (2015) mostly deal with implicit, not (yet) overt disagreement about words and word use.¹ Such implicit disagreements can be more confusing because the borderline between metalinguistic and world-level disputes² is even less clear in those cases.

There are two ways to understand the “implicit” versus “explicit” distinction in this context. On the one hand, one might consider any statement that quotes linguistic material as explicitly metalinguistic. For example, “The word ‘athlete’ only applies to humans” is such a statement. This distinction is not very useful, though, because almost the same can be expressed without quoting linguistic material by saying “Only humans can be athletes”, “To be an athlete requires to be a human”, or simply “Athletes are human.”³

A better, though less precise way of drawing the distinction is by considering what speakers are aware of. If the speakers are aware that they disagree with each other about the meaning of a term or how it is or should be used, then the metalinguistic dispute is explicit. Otherwise, the dispute is implicit. As Plunkett and Sundell (2013) argue, metalinguistic disputes are often implicit in this sense. Often discourse participants are not fully aware that the disagreement concerns the use of a term or its meaning. For instance, in the much-discussed athlete example the dispute indirectly concerns the word “athlete” by addressing what it means to be an athlete (or, what should count as an athlete), yet the discourse participants might not be aware of that dependence on the word.⁴ A puzzling fact about these implicit metalinguistic disputes is that they often start as a seemingly factual dispute and later turn into an explicit metalinguistic dispute. Plunkett and Sundell’s central point is that these kinds of disputes are not merely about words but can also be substantial.

The theory theory is mainly concerned with explicit and implicit *normative* metalinguistic disputes. Descriptive metalinguistic disputes only play a minor role as a particular case. The normativity in question is very weak, however. When an honest and sincere speaker argues on the basis of some (usually hidden) theory or world view, those arguments presuppose that if the hearers disagree with the proponent’s position, then they are making a mistake. In this very broad sense of “normative”, the sense of making a mistake if you do not follow a rule, every rational discussion between honest and sincere speakers is normative.⁵ What Plunkett and Sundell call descriptive

1 Thanks to an anonymous reviewer for pointing this out.

2 For lack of a better term, the adjective ‘world-level’ is at this time artificially constricted not to include theories about language. Cappelen (2018) uses it similarly.

3 In terms of truth-conditions and embedding into propositional attitudes, any explicitly quoting variant differs from a non-quoting paraphrase because it introduces a dependence on the respective language. However, these differences do not have a substantial bearing on a dispute, which always requires exchanging utterances in a given language.

4 The example is from Ludlow (2008, p. 118), compare Ludlow (2014, p. 78).

5 Burgess and Plunkett (2013 ab) and Plunkett and Sundell (2013) sometimes seem to have this weak normativity in mind, but they often also talk about social norms and moral normativity, for instance, when they discuss the waterboarding example. From the present point of view, the latter type of dispute is either empirical-descriptive, if they merely resort uncritically to the existing social norms, or based on *moral theories*. Work in metaethics such as Laskowski and Finlay (2018) is also concerned with the stronger normativity of moral theories that is not meant with ‘normative’ in what follows. From the present perspective, moral theories are merely particular kinds of theories with strongly normative (deontic) implications, as opposed to, say, theories in the natural sciences from which empirical predictions can be derived. This view seems to be compatible with recent remarks by Plunkett and Sundell (2019, pp. 17–21) in reply to Cappelen (2018). However, a worry remains that they do not distinguish

metalinguistic disputes are then disputes that are based on a theory about common word usage or commonly-conceived “everyday” language meaning. For the theory theory nothing else is special about these disputes, and in the end, all metalinguistic disputes are weakly normative.

However, these characterizations are still rough and one might ask whether it is possible to clearly distinguish metalinguistic from normal disputes. According to the indirect meaning characterization thesis, which will be laid out in the next section, the answer to this question is negative. Every theory constrains the range of possible explicit definitions one might give to a term, and so the meaning of terms is characterized indirectly by a theory whenever that theory is considered, as long as the respective term is not defined explicitly. Hence, the distinction between metalinguistic and other types of disputes can principally not be sharp. Every metalinguistic dispute is based on a world-level dispute.

In many of their joint and individual publications Burgess, Plunket, and Sundell suggest similar views about the fuzziness of the borderline between metalinguistic and substantive disputes. For example, they often talk about terms and concepts interchangeably, and this only makes sense if disagreements about the meaning of terms go hand in hand with disagreements about conceptual systems. The theory theory is compatible with their position in that respect. It puts the main emphasis on the underlying theories, whereas they emphasize the normative aspects of metalinguistic disputes and the fact that they often concern the social role of terms. However, according to the theory theory a term’s social role is just one among many factors that go into the background theory on which those disputes are based. Moreover, a term’s social role is somehow given within a speaker community. There is no reason to believe that the social role can vary from speaker to speaker. In contrast to this, theory dependence is more fine-grained. Theories and opinions indirectly characterize meanings when individual speakers disagree with each other, and the social role of a term alone does not explain this theory dependence.

Laskowski & Finlay (2018) have recently criticized Plunkett and Sundell’s approach for overemphasizing the role of metalinguistic disputes about (strongly) normative terms. In their view, these disputes are usually substantive. Instead, they suggest a linguistic “neo-classical” concept analysis of these terms according to which theorists make abductive inferences to the best account of what these might mean. From the perspective of the theory theory this approach is not very promising. If the indirect meaning characterization thesis is right, then people with sufficiently different background beliefs and theories (in a broad sense) will characterize the same term differently. Hence, there *are* metalinguistic disputes but at the same time these are substantive disputes. In the end, the disagreements are always about the merits of underlying theories, and can therefore not be investigated, let alone be decided, based on concepts, meanings, or social roles that are already shared and more or less uncontroversial within a speaker community.

There is recent work in metaphysics by Thomasson (2015, 2017) that also needs to be mentioned. Her deflationary approach to ontology is also based on theories, and she characterizes metaphysical disagreements as normative disputes about conceptual choices. However, she explicitly rejects taking theoretical virtues as the main (anti-realist) criteria for evaluating the merits of theories and instead argues that theoretical virtues focus on empirical adequacy and that the remaining virtues like elegance, connectivity with other theories, or internal coherence do not suffice for clearly determining the merits of metaphysical theories (Thomasson, 2017, pp. 368-370).

It is hard to see why this should be taken as an argument for ontological deflationism rather than as an argument against certain kinds of metaphysics, but we can put this matter aside. The theory theory is not concerned with the adequacy of theoretical virtues for deciding metaphysical

clearly enough between relying on normative moral theories and the much weaker general normativity of metalinguistic disputes.

quandaries. It merely acknowledges that we use theoretical virtues to compare and evaluate competing theories. Neither does it presume that every metalinguistic dispute is substantive, nor that the world-level theories behind them are always good, nor that theoretical virtues can decide every dispute. This more modest use of theoretical virtues remains compatible with Thomasson's more ambitious theses about metaphysics without endorsing them.

2.3 Indirect meaning characterization

According to the *indirect meaning characterization thesis*, whenever a term is not explicitly defined, a sentence in which the term occurs will indirectly characterize the meaning of that term, as long as the term is the subject of a predication or the implicit or explicit subject of quantification.

For example, suppose that someone is convinced that atoms are the smallest indivisible building blocks of nature.⁶ Call this Theory A. If this part of chemistry is true, then it would be possible to make the attribute *indivisible* part of an explicit definition of the meaning of "atom." If it is not made part of an explicit definition, then the law-like statement that atoms are the smallest indivisible building blocks of nature indirectly characterize the meaning of "atom." To see this, suppose Mary knows nothing about atoms and hears "Atoms are the smallest indivisible building blocks of nature." If she considers the testimony reliable and epistemically trustworthy, then she is justified in forming the belief that "atom" means "smallest indivisible building blocks of nature." Of course, if John, who counts as a notoriously unreliable and untrustworthy source of information, utters this sentence, then she might also process the information differently, keeping track of that unconvincing theory of atoms without endorsing it. In this case, the right way to extract metalinguistic information from the utterance would be to extract the proposition that *according to John*, "atom" means "smallest indivisible building blocks of nature", but that this definition is probably inadequate and his theory of atoms likely lacks the theoretical virtues to count as an acceptable theory.

Suppose we conduct experiments and learn that atoms can be split. According to the new theory, call it Theory B, atoms can be split. They are not indivisible, hence also not the smallest indivisible building blocks of nature. Instead, they are the smallest building blocks of nature with the characteristic properties of chemical elements. When we change from A to B, the attribute *indivisible* can no longer be used in a definition of "atom", as the law-like statements of B characterize atoms in a way that is incompatible with the law-like statements about atoms in A. Suppose now, contrary to the original assumption, that "atom" is *not* defined explicitly in those theories. Then the theories still indirectly characterize the meaning of "atom" in different, mutually incompatible ways. They do so independently of the question of whether "indivisible" *should* be used to define "atom" in Theory A or not; different sets of law-like statements concerning atoms in A and B constrain the range of possible definitions of the term. That is a case of *indirect meaning characterization*.

It helps to think about the connection between theories and meaning in terms of what an agent can learn about the entities under consideration from a theory if that theory was free of contradictions, fruitful, and also possessed all other theoretical virtues one might imagine.⁷ In such an ideal and fully acceptable account of an aspect of reality, some terms will be defined explicitly and others may be characterized in the way outlined above, based on an existing foundational common sense ontology and a fixed logical and mathematical vocabulary. We need not address the

6 The term "atom" is used in this way in mereology. For the sake of the argument, we may treat the distinction between atoms in mereology and atoms in physics as a case of ambiguity in different jargons and ignore the use of the word in mereology in what follows. A and B are theories of the physical world.

7 See Keas (2018) for a recent systematization of theoretical virtues. Which particular taxonomy of theoretical virtues is chosen does not matter in what follows, though.

question of whether two *ideal, yet mutually incompatible theories* A and B could be about the same topic. This problem can be left for another occasion. Instead, let us address the easier to answer question of what the compatibility of theories in general, or lack thereof, implies for disagreements about the meanings of terms in those theories.

Here is an approximation that turns out to be incorrect at a closer look and will be refined in Section 2.7. Suppose there are two ideal theories A and B and consider a general term or predicate t used in both of them. In this case, if we rename t as t_A in A, t as t_B in B, and the theories remain compatible with each other after renaming, then a disagreement about t based on A and B would *not* be substantive, even if t_A and t_B were not defined explicitly. Whatever an agent learns from statements containing t_A will be compatible with what the agent may learn from statements containing t_B . Chalmers (2011) provides a similar criterion. On his account, “A dispute over S is (broadly) verbal when, for some expression T in S, the parties disagree about the meaning of T, and the dispute over S arises wholly in virtue of this disagreement regarding T” (Chalmers, 2011, p. 524). If a dispute about t in a sentence S can be resolved in the above way by renaming, then it was wholly in virtue of t , hence according to Chalmers’s criterion, it is verbal.

The problem is that in practice it is rarely the case that two theories *about the same topic* are fully compatible with each other (if that is even possible), and if they are not about the same topic, then a dispute about t in the above example amounts to some form of talking past each other. Expanding on the definition given in the introduction, we will see in Section 2.6 that such disputes may still be substantive, but before this question can be addressed more has to be said about topic continuity.

2.4 The problem of topic continuity

In laying out his diachronic theory of *conceptual engineering*, Cappelen (2018) spends a lot of effort on explaining why topic continuity can be warranted in his framework. Translated into the theory theory, the problem is this: How we can know that theory A and theory B are about the same topic? How do we know, for instance, that theory A in which atoms are indivisible and theory B in which atoms are not indivisible talk about atoms in the first place?

Cappelen’s attempt at a solution is based on controversial theses. He resorts to the *lack of control thesis* and the inscrutability of extensions on the one hand, and rough equality between extensions and intensions on the other hand. However, as Sawyer (2020) argues, rough equality of extensions and intensions does not work for a genuine meaning *replacement*. Consider for instance the pseudo-biological account of a race during the 3rd Reich in Nazi Germany. It was based on a vast number of scientifically false claims and plays no role in modern biology.

In contrast to this, modern social theories of race are based on self-identification or on social roles, or on a combination thereof, and the idea that oppressed groups may need to use a principally defective vocabulary to remain identifiable as a group that is oppressed or to be able to identify the oppressors. These two types of theories, the pseudo-biological race ideology and a (broadly-conceived) social role theory of race, do not have anything in common, or so it seems at first glance.⁸ One has replaced the other. Semantic amelioration in the sense of Haslanger is even explicitly designed to disrupt oppressive practices by attempting to substitute one defective meaning with another meaning that makes the oppression explicit, thereby leading to a change of social practices that in turn changes the corresponding social constructions (Haslanger, 2012, pp. 20–21).

⁸ There can be many different social role theories, which may be or may not be combined with self-identification to more or less of an extent. To talk about this as one theory is thus a simplification for the sake of argument.

In all of these cases, one meaning associated with a term is replaced by another meaning. However, as Cappelen lays out, Strawson's (1963) reply to Carnap (1950) seems to question such a replacement's meaningfulness. Strawson's accusation is that the new theory talks about something else, that the topic was changed. If in Carnap's account the topic already changes with every fruitful scientific explication of a term, then *a fortiori* the topic would change in the above cases of a full replacement.

So how does the theory theory handle such perceived topic changes? I would like to argue for two theses in the following paragraphs: First, all interesting examples of topic continuity can be explained by measurement operations that pick out roughly the same external entities in fallible ways. Second, even when measurement operations do not warrant topic continuity in a particular case, the dispute may still be substantive as long as the underlying theories are mutually incompatible or interesting enough. If that is correct, then topic continuity is not a necessity, but rather the cherry on top of the icing of a theory theory account of metalinguistic disagreement.

2.5 Measurement operations and topic continuity

Measurement operations were discussed extensively after Bridgman (1927) laid out the foundations of operationalism in the philosophy of science. Operationalism was influential in psychology and one of the driving forces behind essential developments in *measurement theory*. However, Bridgman's program as a whole was doubted from the start by physicists and no longer plays a role in contemporary physics because it put a too strong focus on measurement. Theoretical notions have turned out to be crucial in physics and not every useful notion can be grounded in measurement. Luckily, full-fledged operationalism need not be assumed to get the theory theory going. It suffices to borrow the term *measurement operation* from Bridgman's program without assuming any reductionism about theoretical terms.⁹

In what follows, a *measurement operation of a term α* is understood as any method that is commonly associated with the term α in the respective speaker community and allows speakers to fallibly and roughly identify or measure particulars falling under that term. Understood in this way, measurement operations may take part in a description of the meaning of a general term, but they do not need to. They may be merely associated with the term as part of a common practice or everyday meaning that an expert would not include in a definition.¹⁰ Since any measurement is fallible, the measurement operation associated with the term is also fallible. Moreover, not every speaker needs to associate the same operation with a given term to warrant topic continuity. Instead, the thesis is more modest: If a general term α has one or more associated operations P_1 to P_n , for roughly the same set of objects $\{a_1, a_2, \dots, a_k\}$, then these operations may warrant topic continuity even if the meaning of α changes. They do so by fallibly identifying those objects as particulars that presumably fall under whatever α means. Moreover, in the current context "roughly" means whatever theory of rough equality between sets is adequate. An account based on Zadeh (1965) would probably do, but the details can be left open. It suffices for current purposes to say that P_1 to P_n need to pick out roughly, but not exactly, the same sets, and that different speakers may associate different measurement operations from $\{P_1, P_2, \dots, P_n\}$ with the same term.

9 Although he was accused of it, Bridgman himself was not a reductionist about theoretical entities. These are discussed in the section 'Models and Constructs' (Bridgman, 1927, pp. 52–60), where he writes: 'The moral of all this is that constructs are most useful and even unavoidable things, but that they may have great dangers, and that a careful critique may be necessary to avoid reading into them implications for which there is no warrant in experience' (Bridgman, 1927, p. 60).

10 Hereby, "associated" is meant in a psychological sense very similar to associative meaning in Leech (1974), and it is important to bear in mind in what follows that an operation need not be part of the meaning of a term.

This may look like the old “loose bundle view” of reference that Kripke (1972), Putnam (1975), Burge (1979), and so many others have rejected, but it is not. Operations in the above sense need not be part of the meaning of a term. Instead, they are *used* by speakers to putatively identify the particulars that fall under a specific term. Two speakers, say, John and Mary, may associate two different operations P_1 and P_2 with a term α and develop mutually incompatible α -theories A and B. They will still talk about roughly the same topic if P_1 and P_2 can be used to identify or measure roughly the same sets of particulars. Theories A and B could be utterly mistaken, John and Mary’s definitions of “ α ” woefully inadequate and contradictory – their theories will still be about the same topic if P_1 and P_2 pick out roughly the same set of particulars.

2.6 Examples of topic continuity

The first example involves explicit definitions. It provides a good starting point because examples based on explicit definitions tend to be less controversial than examples based on implicit meaning characterization. As is well-known, the Urmeter was used to define one meter. Later the definition was based on the wavelength of light emitted from a Krypton-86 source, and finally, one meter was defined as the length that light travels in a vacuum in a 1,299,792,458th of a second. These explicit definitions are about the same topic because they are based on different ways to measure roughly the same length. When the Urmeter was the norm by convention, the distance the light travels in 1/1,299,792,458 seconds would have been slightly off if it had been measured very precisely. Likewise, nowadays the Urmeter may be a little bit off according to the new convention. In all of these cases, the actual measuring serves as an operationalization of the term “meter”, which happens to be defined by convention.

Turning to less arbitrary definitions, consider the general term “atom” again and a definition in Theory A that presumes atoms are indivisible. In Theory B atoms are divisible and, if they are defined rather than just being characterized indirectly, they are defined very differently from Theory A. Nevertheless, there are practices in chemistry of weighing molecules and computing the combinations of different elements that allow one to identify and measure them in a compound. Taken together, these practices of identifying chemical elements—that is, different kinds of atoms—and their properties are operations that warrant the topic continuity between A and B.

Turning to more contested cases, consider possible commonalities between the pseudo-biological race ideology of the Nazis and modern conceptions of race-based on identity and social properties of groups. It is far less clear whether topic continuity holds in this case. Topic continuity between an alleged “Jewish race” stipulated by the Nazis and contemporary Jewish identity is based on religious affiliation and the historical connection that much of the modern identity was caused and influenced by the Holocaust, but certainly not on any physical features that the Nazis attributed to Jews.¹¹ However, things may look different for race talk about “black” and “whit” people. On the one hand, there is the historical connection that many oppressors of black people have and are still attempting to justify their racism by ill-conceived pseudo-biology. On the other hand, there are also shared operations in place that warrant topic continuity, at least to some extent. Regardless of how justified or unjustified this way of talking is (it seems to have no meaningful biological basis either), at least in some areas of the world talk about “white” and “black” people seems to be based on skin color and other physical traits that count as measurement operations.¹² So it seems that at

11 There was strictly speaking no “Jewish race” according to Nazi doctrine because they were supposed to be a mix of many different “inferior races.” The underlying pseudo-science was highly inconsistent. While on the one hand the Nazis often claimed that Jews could be identified by their physical traits, on the other hand they also required an *Ariernachweis* (Aryan certificate) based on a genealogical analysis and religious affiliation. The reason for this inconsistency is that the distinctions were ultimately ideological and political.

12 See also Haslanger (2012: p. 306).

least some of the operations associated with “human races” have carried over from the pseudo-biological ideology to modern and widely accepted ways of (self-)identifying groups partly based on their physical attributes, and hence at least for these parts of race talk there is topic continuity. This ought not to be surprising. Historically speaking, many groups have been oppressed based on abhorrent and scientifically untenable theories and world views, and therefore have been forced to use these deficient theories and world views to identify themselves as an oppressed and disadvantaged group. One of the perversions of actual oppression is that it may force oppressed groups into making differences they would not find meaningful under better circumstances, had history turned out another way.

Other cases that seem to be partly based on operations and partly subject to a continuous topic change are words like “gender”, “family”, “nation”, “cool”, and “faithful.” There are good reasons for believing that the meanings of these words have changed over time and that their denotations may be socially constructed as opposed to being social kinds similar to natural kinds. In spite of that, their topic continuity is warranted by measurement operations at least to some extent. For example, typical and functional families usually live together most of the time, provide care and love for their members (their children in particular), consist of parents and children, maybe also include grandparents, and so forth. These properties manifest themselves in various visible ways that indicate a family’s existence, and even if one of these properties is taken away as a criterion for a family, the others remain. So if a family has traditionally been regarded as including a father and a mother, for example, then the extension of this concept to additionally include the combinations father+father and mother+mother are not substantially different from other changes, such as a transition from large families to small families or the transition from the stereotypical family with father and mother to a family consisting of a single mother and her children. The associated measurement operations serve as soft criteria for determining topic continuity.

2.7 Topic continuity is not a necessary condition for a substantive dispute

In all of the above examples, one or more measurement operations are associated with the term in question, although these need not be part of its definition or perceived meaning, and their application warrants topic continuity. The process of measuring what different theories talk about is never perfect, always fallible, and only rough equality between the sets of objects picked out by those measurements is needed to warrant topic continuity. In that respect, the suggestion is very similar to what Cappelen (2018, pp. 109-117) proposes. That being said, according to the theory of metalinguistic disagreement, topic continuity is not a necessary condition for making a metalinguistic disagreement substantive, and lack of topic continuity does not necessarily indicate that it is pointless or merely verbal.

Why not? The reason is, simply put, that in the end, we never argue about the meaning of terms in isolation, but always based on background theories. These theories can be more or less adequate and have more or fewer theoretical virtues. Therefore, a dispute about those theories can be substantive and meaningful even if there is no topic continuity between them. Even if we do not know whether Theory A and Theory B are really about the same topic, one of them might be much better than the other in terms of theoretical virtues, giving us ample reason to accept one and reject the other, whether they roughly agree on the topic or not.

For instance, it is tough to determine whether esoteric theories and ideologies like “diamond therapy” and “levitated water after Hacheney” have any topic in common with medical and physical theories. At the same time, it may be obvious that the former have no scientific merits. The question of topic continuity only becomes relevant once theories are compared that are roughly on a par in

terms of their theoretical virtues, not when one or both of them can be refuted easily by an unbiased and informed assessor.

This is not to say that topic continuity is wholly optional. Within the acceptable range of theories and opinions, when there are arguments for and against them, topic continuity warranted by suitable measurement operations is vital to rule out misunderstandings and talking at cross purposes. The following cases can be distinguished:

1. If two theories are compatible with each other and there is no topic continuity between them, then accepting one should have no bearing on the other and vice versa. Disputing terms used in one theory from other theory's perspective may then lead to talking at cross purposes. However, the dispute will remain substantive if both interlocutors have a vested interest in weighing the pros and cons of those theories despite potential confusion about the topic, if the theories are important.
2. If two theories are not compatible with each other and there is topic continuity between them, one has to prefer one theory over the other or remain agnostic about the choice, based on the virtues of the theories, and cannot fully endorse both of them at the same time.
3. If two theories are not compatible with each other and there is no topic continuity, then this may indicate fundamental problems with the theorizing and even incoherent world views. This may lead to a choice between the theories based on their virtues, but may also trigger a wider investigation as to why they are incompatible with each other. Conflicts between General Relativity Theory and Quantum Mechanics in contemporary physics provide an example.
4. Finally, if two theories are compatible with each other and there is topic continuity, then they seem to highlight different aspects or levels of description of the same subject matter and may both be accepted or rejected based on their theoretical virtues. Metalinguistic disputes between proponents of such theories are also a form of talking at cross purposes, though still potentially substantive if they clarify or improve a theory or viewpoint.

The fact that metalinguistic disputes are based on underlying theories (in the broad sense of the word) can thus be taken as a justification for Plunkett and Sundell's (2013) claim that metalinguistic disputes are often substantive. Notice that only case 2 implies a lack of co-tenability. Moreover, only cases 1 and 4 may lead to an insubstantive or "merely" verbal dispute, and they do so only if the theories are not of particular interest to the interlocutors, if nothing important hinges on them and no important social roles are connected with the terminological choice under consideration, and if the dispute can be resolved by refining the terms. For example, a disagreement about "Tomatoes are fruits" can be considered an instance of case 1 because it can be further discerned into "Tomatoes in the biological sense are fruits" and "In cooking, tomatoes often count as vegetables." This dispute is non-substantive only if biology and cooking are compatible theories about different topics and there is nothing gravely wrong with the "tomatoes in biology" and "tomatoes in cooking" sub-theories.¹³

13 It could also be taken as an instance of Case 4, though this would be an odd choice. While it is true that "tomatoes in biology" and "tomatoes in cooking" are about tomatoes, these are not natural units of inquiry in this case. The theories in question are biology and cooking, and these are not theories *of* tomatoes in particular. However, the problem of discerning natural units of inquiry is not addressed here. Maybe it does not have a clear solution in general, and for the current purposes, it does not matter whether we describe the example as an instance of Case 1 or Case 4.

In general, however, disputes of type 1 and 4 will often lead to substantive metalinguistic disagreements, and the accusation of “merely” talking at cross purposes is usually not justified. To give an example, suppose Theory A is Wilhelm Reich’s orgone theory and B is the current state of the art in psychological research about human sexuality. This will likely be an instance of case 4, but we may just as well assume case 1 for the sake of argument, that the theories are compatible with each other. In both cases, it would be perfectly fine for Mary to argue, based on Theory B, that “orgone energy” is an ill-conceived notion and that the term ought not to be used in an explanation of certain forms of neurosis. Even a seemingly purely metalinguistic dispute about the meaning of “orgone” is substantive in this context because it is ultimately based on weighing the scientific merits of different theories. Theory A lacks theoretical virtues that B has, and the formulation of Theory B does not require the use of the term “orgone.”

3 OBJECTIONS AND REPLIES

In this section, three objections to the theory theory of metalinguistic disagreement are addressed. Although I find them ultimately unconvincing, they highlight important features of the account that sets it apart from alternatives.

3.1 The relativist objection

One worry is that talking about theories in the loose sense suggested above constitutes a form of contextualism to which standard relativist objections apply. To cut a long story short, the objection goes as follows. If the indirect meaning characterization thesis is correct and two speakers John and Mary presume different background theories A and B that characterize a term α in different ways, then at the level of the semantic content of utterances containing α , even seemingly contradictory statements may remain compatible with each other. Hence, there is no real disagreement and John and Mary talk at cross purposes.¹⁴ This objection can be generalized by not pinning disagreement on semantic content alone, but instead looking at whether those statements can be rationally accepted at once, or by looking at co-tenability and joint reflexive accuracy of attitudes that speakers may hold if they were to endorse the theories fully.¹⁵ The objection is that all these criteria for determining disagreement fail for the theory theory of metalinguistic disagreement.

To fix this problem, one might “go relativist”, that is to stipulate a level of content that is true or false relative to the contextually variable element. This would mean stipulating a level of content that was true or false relative to theories. The original form of moral relativism by Harman (1975) seems to be based on this idea, though later moral relativists like Harman (1996) and Wong (1984, 2006) have defended positions that are closer in spirit to semantic contextualism.

The problem is that a relativist semantics does not yield the intended result for contextual dependencies on whole theories. Suppose John says (1) “Atoms are the smallest indivisible building blocks of nature” and Mary answers: (2) “No, they aren’t. They can be split.” The worst way of explaining such an example would be to stipulate that (1) is true and (2) false relative to John’s theory of atoms, and (1) is false and (2) true relative to Mary’s theory of atoms. Of course they are, but these conditions do not reflect the truth-conditions of the sentences. John’s theory has been

14 Compare Khoo (2017, pp. 257–260).

15 See Plunkett and Sundell (2013, p. 11), MacFarlane (2014, Chapter 5).

falsified and Mary's theory has been confirmed. John's statement happens to be false and Mary's statement happens to be true because that is the way nature is. To say that John's utterance is true relative to *his* theory is to merely restate, in a convoluted way, that John believes it is true. Even though the meaning of terms in (1) may depend on John's theory and Mary needs to keep track of this, it *also* depends on nature insofar as his definition of "atom" is likely inadequate, and his theory does not adequately grasp an aspect of reality.

That is not to say that a relativist account cannot have limited applications at the theory level, at least in principle. If values *only* depend on human nature or do not depend on nature at all, then there may be classes of adequate value theories that are assessor-theory relative in the above sense. However, even if specific forms of assessor-theory relativism were adequate for moral discourse, this would not imply that the same approach is adequate for other types of discourse involving metalinguistic disagreements. Even if one argues that theoretical virtues may be, to some extent, in the eye of the beholder and subjective, this does not suffice to justify general theory-based relativism.¹⁶

Another problem with the relativist critique is that co-tenability and related notions are not well-suited to determine disagreement about theories, at least not if co-tenability is spelled out in terms of a rational agent's ability to have corresponding attitudes such as beliefs and desires at the same time. We are principally capable of understanding conflicting theories, opinions, and world views. We can consider their merits and hold them in our minds without agreeing with them. The same rational agent may not be able to knowingly fully *endorse* two conflicting theories at the same time, but endorsing a theory is not a necessary condition for arguing about it and discussing its merits. We can disagree about theories that we do not endorse at all, because one theory may be better than another while still not meeting one's criteria for believing it to be acceptable and well-confirmed. For example, the idea is flawed that Mary and John disagree in terms of theories if and only if Mary endorses Theory A and John endorses Theory B, and A and B are not co-tenable in the sense that John cannot also endorse A and Mary cannot additionally endorse B without running into some rational conflict like a contradiction or pragmatic incoherence. This criterion is too strong since John and Mary could disagree about the theories even if neither of them endorsed any of them. If, in contrast to this, co-tenability is just meant to be understood in a way that an agent may consider two theories at the same time and contemplate their merits, then lack of co-tenability does not generally indicate disagreement either. There is even a case to be made for the opposite view: For John to be able to disagree with Mary *rationally*, he has to first consider and, at least to a certain extent, understand Theory A and how it conflicts with Theory B.

From the perspective of the theory theory of metalinguistic disagreement, worries about talking at cross purposes do not apply if topic continuity is warranted and only apply to a limited extent if topic continuity is not warranted (cases 1 and 4 in the above list). We are smarter than the relativist critique presumes and can track, to a sufficient degree, other people's theories about various topics, including their definitions and indirect meaning characterizations. This process may fail, and we may sometimes become confused or talk at cross purposes, but that is just the way

16 There is another, more principal worry with theory relativism, that is, any position according to which the truth and falsity of statements or semantic contents is relative to a whole theory. The position is based on two vicious regresses. First, if the truth value of any utterance U is relative to theory A, then the truth value of any sentence in A should also be relative to another theory B, and so forth, ad infinitum. Second, to decide whether the truth value of types of utterance U is relative to theories, a metatheory about the respective domain of inquiry is needed. For example, moral relativism is a particular *metaethical theory*. But if global theory relativism is true, then that metatheory is itself relative to a meta-metatheory, and so forth. I believe that both regresses are a problem for the position, but do not want to delve into this discussion here any further. They distract from the real issue, which is that there is no good evidence for *general* theory relativism in the first place.

some conversations go; of course, a theory of metalinguistic disagreement must leave room for such cases.

3.2 Externalist worries

Perhaps the strongest objection to the approach presented so far comes from certain varieties of semantic externalism. An externalist might complain that the theory theory is based on the hidden assumption that the meaning of general terms and predicates is composed out of logical combinations of other, less complex meanings of predicates and terms, and that this view about lexical meaning is incompatible with Kripke's and Putnam's arguments for externalism. Instead, whatever falls under a general term is decided by its extension, which was, in turn, fixed in a broadly conceived indexical way. "Water" refers to ionized H₂O plus some impurities in our speaker community, because water *is* H₂O. If water was XYZ where we live, then "water" referred to XYZ. All of this is true independently of the properties we believe constitute water or define "water" Even before humans had a clue about chemistry, "water" referred to H₂O, because we use the term indexically to pick out a natural kind. That is the lesson from Kripke (1972) and the first part of Putnam (1975). We may call this position *indexical externalism*.

Before addressing this type of objection in more detail, it is useful to distinguish this position from another version of externalism. According to *social externalism* of Burge (1979), what falls under a general term is ultimately decided by experts in the speaker community and speakers need not know the experts' criteria to use a general term competently. This version of externalism is not problematic for the theory theory, it is in fact implied by it. Some theories are scientific, and it is common to defer terminological disputes to those who develop scientific theories about a subject matter. The experts, in turn, often disagree about the most adequate meaning of a term or predicate, what role it plays in the theory as a whole, and everything that has been said above applies to these scientific disputes as well.

What about indexical externalism then? One answer can be found in Plunkett and Sundell (2013, pp. 26–28). They argue convincingly that normative metalinguistic disputes can occur between speakers who agree on all relevant facts. For instance, the speakers in Ludlow's "Secretariat is an athlete" case may agree on all relevant facts about the situation like Secretariat's achievements, how long the horse has been trained, how many races it has won, and so forth. So the dispute cannot be about these facts. Nevertheless, a normative metalinguistic dispute about the question of whether horses can be athletes may make perfect sense and a debate may be important in certain contexts, for instance, with respect to animal rights.

However, there is a problem with indexical externalism that goes beyond the cautious reply in Plunkett and Sundell (2013). Indexical externalism is an overall implausible metasemantic view if it is meant to apply to *all* general terms and predicates. Tigers have roughly the same genetic makeup and water consists roughly of partly ionized H₂O. There is a solid naturalistic basis for assuming that "tiger" and "water" are natural kind terms. In contrast to this, there is no solid naturalistic basis for biological race talk; hence, there is also no natural kind that corresponds to the respective terms. To give another example, as Einstein and many others realized after the Michelson-Morley experiment, nothing corresponds to "aether" in reality. In spite of that, disputes about the nature of aether and the meaning of "aether" took place and were without doubt substantive, meaningful, and worth having. Since there *is* no aether, an indexical externalist cannot point to it to fix the extension of "aether." At the same time, it is bad lexical semantics to claim that the meaning of a term like "aether" cannot be decomposed into logical combinations of predicates

and other terms. It clearly can be semantically decomposed within the theory of aether, and this decomposition plays a crucial role in the falsification of the theory.¹⁷

Modern externalists will not consider this criticism very strong, as they acknowledge that only the meaning of *some* general terms is fixed indexically. Other terms, maybe most, are based on external facts, yet not in the same way in which natural kind terms hinge on their extension and the environment of the linguistic community. According to Haslanger (2012, pp. 132–136), for instance, many terms that are frequently under dispute are socially constructed. Her “debunking constructionism” remains compatible with externalism because these constructions can be taken to be based on the question of whether the best available social theory finds them valuable and useful. Moreover, the meaning of terms can be changed by linguistic actions like defending ameliorative definitions and by social actions, and so there is ample room for normative metalinguistic disputes. Like in the theory theory, in her account the borderline between a metalinguistic and a world-level dispute is vague and overall not important. Changing linguistic usage and definitions will change social reality to some extent and, vice versa, changes of social reality will often effect meaning changes. This type of externalism remains compatible with the theory theory. In fact, a proposal for a certain semantic amelioration is a normative theory, in the broad sense of “theory” used here, and should be evaluated according to its overall merits.

In contrast to this, Cappelen (2018) has defended a form of externalism that seems to be incompatible with the theory theory at first sight. Among other elements, it is based on the *lack of control thesis*. According to this thesis, we do not have much control over the linguistic meaning of expressions since it results from various external processes whose workings we do not even fully understand. To these belong views from indexical externalism such as dubbings and causal chains, as well as more pragmatist elements like patterns of language use. Since we lack control over these factors, normative metalinguistic disputes are mostly futile. Words have their meanings because of a long history of uses in a shared external environment.

There is something wrong with the lack of control thesis. It may look convincing for examples of attempting to intentionally change language in everyday conversations but starts to look reasonably implausible when more elaborate conversations are considered, for example conversations in scientific contexts. An incorrect definition of “atom” that involves being indivisible can be corrected when a better theory is proposed, and even school kids nowadays know that “aether” as a medium for light and the planet “Vulcan” between Mercury and the Sun do not exist. So there must be a certain amount of control.

However, suppose that Cappelen is right about our lack of control. Is this view incompatible with the theory theory? It seems clear that it is compatible. The theory theory states that if we want to resolve a dispute about words, it matters not only what you claim but also how you *justify* those claims and which theory you promote by justifying them. To resolve a disagreement about the adequate definition of “atom”, for instance, physicists cannot merely claim that atoms are divisible, they have to contrast the theory as a whole with existing theories, explain why it is better confirmed by existing evidence, is simpler or mathematically more beautiful, is more economical than the other theory, has more fruitful implications for other theories, and so on. The same would hold for “family” even if there was a social kind for it akin to a natural kind.

This view is not only compatible with Cappelen’s externalism, it even supports it to some extent. One of Cappelen’s points is that supposedly metalinguistic disputes are in reality world-level

17 Closely related to indexical externalism is primitivism, according to which certain expressions or concepts cannot be further analyzed. Moore (1903) defended this for “good”. Laskowsky and Finlay (2018, pp. 539–540, 546–550) point out problems with this view and I have criticized it in Rast (2017). In a nutshell, my critique is similar to the above one, that primitivism does not invalidate the fundamental theory dependence of our justifications upon which rational disagreements rest.

disputes. The theory theory states the same by pointing out that substantive metalinguistic disputes, those that are not just about words, are based on potentially conflicting theories and opinions about the world. Since Cappelen does not deny that certain disputes appear to be metalinguistic, for example those that are overtly about the meaning of a word, the two approaches complement each other. According to Cappelen, metalinguistic disputes are world-level disputes in disguise. According to the theory theory, metalinguistic disputes are based on theories about what constitutes the right meaning of a certain word, but for such a theory to make sense it needs to be supported by an underlying world-level theory.

So there is a difference, but it is not as big as it might seem at first glance. The same can be said about indexical externalism. Even an indexicalist externalist who stipulates all sorts of natural and social kinds should endorse the theory theory of metalinguistic disagreement as an account of meaning disputes. The theory theory of metalinguistic disagreement avoids a theoretical commitment to such strong forms of externalism while remaining compatible with them. Whether there is lack of control of meaning change is *their* problem.

3.3 Objections based on holism

There is an interesting objection to the theory theory based on global holism in a broadly conceived Davidsonian setting.

Every theory depends on some background ontology, which involves the meanings and relations between meanings of everyday expressions. To these belong, for instance, expressions like “and”, “or”, “liquid”, “gas”, “object”, “relation”, “property”, “lake”, “meter”, “second”, “prime number”, and so forth. Theories as world views or sets of beliefs will additionally presume a large common-sense ontology with defeasible knowledge like “birds usually can fly”, “birds usually have feathers”, “normal cars have four wheels”, “countries usually have a president”, and so on. When a theory changes, then it may seem as if not only the meanings of terms in some isolated statements of that theory are indirectly characterized, but that the meaning of *any* term of the language is affected by such a change. The idea behind this thesis is that every use of a word such as “and” may indirectly characterize its meaning. Asymmetric temporal interpretations of “and” exist for that reason. There is a difference between “He grabbed the bottle and took a sip” and “He took a sip and grabbed the bottle”, because of the way prior theories (stories, in this case) have indirectly characterized the meaning of “and” as suggesting or hinting at a temporal ordering of its conjuncts. Or so, one might suggest.

A global holist may agree with the theory theory and authors like Ludlow (2014) that there is indirect meaning characterization, but will not accept the locality thesis. Every meaning of every expression results from an interplay of the meaning of any other expressions with which it is collocated, not just in the same sentence or same sets of sentences (local theory), but in the whole history of a person’s uses of expressions of that language. Consequently, no two persons believe exactly the same and for no two persons does the same expression have exactly the same meaning. We instead approximate our meanings to one another by assuming that we are speaking the same language and by applying the *principle of charity* (Davidson, 1967) according to which most of what our interlocutors believe is true. We roughly have the same background ontologies because we have been socialized in similar ways and language is used for communication, but these are only *roughly* the same. Neither can theories be regarded in isolation from the rest of our common-sense ontologies and everyday vocabulary, nor can the effects of theory change on the expressions’ meanings be regarded as local. This is, in a nutshell, the holist critique of the locality assumption of Section 2.1.

My reply to this objection is that the meaning characterization it presupposes seems to be psychological rather than logical, and that, logically speaking, only local holism follows from meaning change within a theory. Global holism is not plausible from a logical perspective.

Psychological meaning holism involves some claim, in one form or another, about associations that speakers make between meanings in their head and terms of a language. Since they can be investigated empirically, there is no doubt that such associations exist and that they constitute conceptual relationships between meanings. Prototypes and stereotypes are based on them, for instance. When such an approach is taken as the basis for holism, it may make sense to claim that every use of an expression may somehow influence the meanings associated with collocated expressions and that more distal global meaning changes may also occur. Such a theory may, for example, explain asymmetric readings of “and”, or the fact that a speaker might be more likely to cluster the word “sin” together with the word “bible” than with the word “calculus.”

In contrast to this, a logical perspective on indirect meaning characterization does not imply global holism without further assumptions. First of all, from a purely logical point of view, a consistent set of formulas may be dividable into subsets that are logically independent of each other. In a broader sense, this is also true of parts of mathematics. For example, as is well-known, if Zermelo-Fraenkel set theory is consistent, then the *axiom of choice* and the *continuum hypothesis* are logically independent of its axioms in a precise and well-defined way. From an even broader perspective, rules and law-like statements about entities in a theory may indirectly characterize some, but not other terms that are used to relate antecedent and succedent conditions. For example, the meanings of “electron”, “proton”, and “neutron” are related to the meaning of “atom” in the statement “Atoms consist of electrons, protons, and neutrons.” However, even though it may be psychologically plausible to assume that the statement also minimally changes the associated meanings of “consist of” and “and” in some subtle ways by usage, the sentence neither predicates nor implicitly quantifies over the extensions or intensions of these expressions and therefore they are not indirectly characterized by the statement. From a logical point of view, a meaning change does not automatically have global effects.

4 SUMMARY

The theory theory of metalinguistic disputes explains why many, if not most, metalinguistic disagreements are substantial and worth having, as Plunkett and Sundell have argued, but without resorting primarily to the social role of terms. Instead, it emphasizes that any disagreement, including metalinguistic disagreements, is based on a conflict between theories—understood in a loose sense as including world views, opinions, scientific theories, stories, and sets of beliefs. Topic continuity can be explained by measurement operations that roughly pick out the same set of entities. If there is topic continuity between two theories, then this suffices to safeguard a metalinguistic dispute against accusations of talking past each other. However, since ultimately the virtues and merits of theories are at stake, metalinguistic disputes are often substantive and worth having even when there is no or only partial topic continuity between the interlocutors’ background theories.

Much of what has been said in this article is also true for ordinary disagreement. Disagreement is never just about isolated facts. If the theory theory is correct, then there can also be talking at cross purposes in ordinary, content-based disagreements. This prediction matches our everyday experience. Arguing based on theories about vastly different topics leads to some form of

misunderstanding that is perhaps best described as a *deficiency of mutual understanding*. Like in the case of metalinguistic disputes, such disagreements may nevertheless turn out to be substantive and worth having if at least one of the respective theories has substantial theoretical merits and describes an aspect of reality that is important to us.

ACKNOWLEDGEMENTS

Many thanks to Pedro Abreu, Bruno Jacinto, Marcin Lewinski, Peter Ludlow, and other participants of the MERELY 2019 conference, as well as the members of the ArgLab Value Seminar at the Nova Institute of Philosophy of the New University of Lisbon for valuable comments on an earlier version of this article and a corresponding talk. A special thanks goes to Ernest Lepore, whose insightful and critical remarks helped me better understand externalism and my own position. I am also grateful to the Portuguese Foundation for Science and Technology for their continuing support.

REFERENCES

- Barker, C. (2002). The dynamics of vagueness. *Linguistics and Philosophy*, 25, 1–36.
- Bridgman, P.W. (1927). *The logic of modern physics*. New York, NY: Macmillan.
- Burge, T. (1979). Individualism and the mental. In French, P. A., Uehling, T. E. & Wettstein, H. K. (Eds.), *Midwest studies in philosophy IV* (pp. 73–121). Minneapolis, MIN: University of Minnesota Press.
- Burgess, A., & Plunkett, D. (2013). Conceptual ethics I. *Philosophy Compass*, 8, 1091–1101.
- Burgess, A., & Plunkett, D. (2013). Conceptual ethics II. *Philosophy Compass*, 8, 1102–1110.
- Carnap, R. (1950). *The logical foundations of probability*. Chicago, IL: University of Chicago Press.
- Cappelen, H. (2018). *Fixing language*. Oxford: Oxford University Press.
- Cappelen, H.: & Lepore, E. (2005): *Insensitive semantics*. Oxford: Blackwell.
- Chalmers, D. (2011). Verbal disputes. *Philosophical Review*, 120, 515-566.
- Davidson, D. (1967). Truth and meaning. *Synthese*, 17(3), 304–323.
- Harman, G. (1996). Moral relativism. In Harman, G. & Thomson, J. (Eds.), *Moral relativism and moral objectivity* (pp. 3–64). Oxford: Blackwell.
- Harman, G. (1975). Moral relativism defended. *Philosophical Review*, 84, 3–23.
- Haslanger, S. (2012). *Resisting reality: social construction and social critique*. Oxford / New York, NY: Oxford University Press.
- Keas, M. N. (2018). Systematizing the theoretical virtues. *Synthese*, 195(6), 2761–2793.
- Khoo, J. (2017). The disagreement challenge to contextualism. In Ichikawa, J. J., (Ed.), *The Routledge handbook of epistemic contextualism* (pp. 257–271). London and New York, NY: D. Reidel.
- Kripke, S.A. (1972). *Naming and necessity*. In Davidson, D., Harman, G. (Eds.), *Semantics of natural language* (pp. 253–355). Dordrecht and Boston, MA: D. Reidel.
- Laskowski, N. & Finlay, S. (2018). Conceptual analysis in metaethics. In McPherson, T. & Plunkett, D. (Eds.), *The Routledge handbook of metaethics*. New York, NY: Routledge, 536–551.
- Leech, G.N. (1974). *Semantics*. London: Pelican Books.

- Ludlow, P. (2008). Cheap contextualism. *Philosophical Issues*, 18, 104–129.
- Ludlow, P. (2014). *Living words*. Oxford: Oxford University Press.
- MacFarlane, J. (2014). *Assessment sensitivity: relative truth and its applications*. Oxford: Oxford University Press.
- Moore, G.E. (1903). *Principia ethica*. Cambridge: Cambridge University Press.
- Plunkett, D. & Sundell, T. (2019). Metalinguistic negotiation and speaker error. *Inquiry*, online first, DOI: 10.1080/0020174X.2019.1610055.
- Plunkett, D. & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, 13(23), 1–37.
- Plunkett, D. (2015). Which concepts should we use? Metalinguistic negotiations and the methodology of philosophy. *Inquiry*, 58(7–8), 828–874.
- Putnam, H. (1975). The meaning of 'meaning.' In Putnam, H. (Ed.), *Mind, language and reality* (pp. 215–271). Cambridge: Cambridge University Press.
- Rast, E. (2017). Value disagreement and two aspects of meaning. *Croatian Journal of Philosophy*, 17(51:3), 399–430.
- Sawyer, S. (2020). The role of concepts in fixing language: a critique of Cappelen. *Canadian Journal of Philosophy*, 50(5), 555–565.
- Strawson, Peter F. (1963). Carnap's views on conceptual systems versus natural languages in analytic philosophy. In Schilpp, P.A. Ed.): *The philosophy of Rudolf Carnap* (pp. 503–18). LaSalle, IL: Open Court.
- Thomasson, A. (2017). Metaphysics and conceptual negotiation. *Philosophical Issues*, 27, 364–382.
- Thomasson, A. (2015). *Ontology made easy*. New York, NY: Oxford University Press.
- Wong, D. B. (2006). *Natural moralities: a defense of pluralistic relativism*. New York, NY: Oxford University Press.
- Wong, D. B. (1984). *Moral relativism*. Berkeley and Los Angeles, CA: University of California Press.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.