13 Externalism, metasemantic contextualism, and self-knowledge

Henry Jackman

Introduction

This chapter examines some of the interactions between holism, contextualism, and externalism, and will argue that an externalist metasemantics that grounds itself in certain plausible assumptions about self-knowledge will also be a *contextualist* metasemantics, and that such a contextualist metasemantics in turn resolves one of the best known problems externalist theories purportedly have with self-knowledge, namely the problem of how the possibility of various sorts of 'switching' cases can appear to undermine the 'transparency' of our thoughts (in particular, our ability to tell, with respect to any two occurrent thoughts, whether they exercise the same or different concepts).

Metasematics, self-interpretation, and contextualism

Philosophers of language can be understood as offering two sorts of semantic theory. On the one hand, they can present an account of what the semantic values of the words in our language *are* and how the values of complex expressions are a function of the semantic values of their parts (what we can simply call a *semantic* theory). On the other hand, they can present an account of how our words came to *get* those particular semantic values (what we can call a *metasemantic* theory).[1] Often, they will present theories of both sorts.

Thanks to Anne Bezuidenhout, Sanford Goldberg. Michael Lynch, Mark McCullaugh, Ram Neta, and Arthur Zucker (as well as audiences in Toronto, Chicago, New Orleans, Barcelona, Lisbon, Paris, and Athens (Ohio)) for comments on earlier versions of this chapter.

[1] Davies (2006) also describes the distinction in these terms, though it is described elsewhere as the distinction between "Descriptive" and "Foundational" semantics (Stalnaker 1997, 1999), "Formal" and "Philosophical" semantics (Brandom 1994), "semantic theories" and "foundational theories of meaning" (Speaks 2011), semantics in the "linguist's" and the "metaphysical" sense (Block 1998), and "formal semantics" and "Philosophical Meaning Theory" (Glüer 2011, 2012).

For instance, within the broadly 'Davidsonian' framework presupposed here one's semantic theory (his "truth theory") understands meaning in terms of truth and satisfaction,[2] with, say, "Chicago" being satisfied by *Chicago* and "city" being satisfied by members of the set of *cities*, and "Chicago is a city" will be true if the object that satisfies the first term is among the set of objects that satisfies the second. In contrast to this 'atomistic' semantic theory, Davidson's metasemantics (which takes the form of his "theory of interpretation") is unapologetically holistic, with the ultimate explanation of how each word gets its meaning being dependent on how other terms in the language get theirs.[3] The Davidsonian understands the satisfaction conditions of any word in a speaker's language as determined by the set of assignments that would maximize the truth of *all* of that speaker's commitments.[4]

Davidson isn't always clear about why one's metasemantics should maximize the amount of truth that the subject believes,[5] but I think that such maximization can be motivated in terms of what will here be referred to as the "Self-Interpretation Principle," namely, when faced with conflicting commitments on an agent's part, we should assign semantic values to his or her words in a way that preserves the truth of the commitments that he or she would hold on to were he or she aware of the tension.[6] So, to take an example from Burge that we will return to later, if a speaker believes both that, say, (1) he has "arthritis" in his thigh, and (2) doctors study how to treat "arthritis" in medical school, then whether his "arthritis" utterances refer to *arthritis* or *tharthritis* (the latter picking out a larger set of ailments including both arthritis and similar pains occurring beyond one's joints) will be determined by which of these two he would treat as mistaken were he made aware of the conflict.[7]

[2] See, for instance, Davidson 1965/1984, 1967/1984.

[3] See, for instance, Davidson 1973/1984, 1974/1984, 1975/1984, 1977/1984, 1979/1984, 1986a.

[4] This is usually explained in terms of the Davidsonian's commitment to the 'Principle of Charity'. For a more complete discussion of this, see Jackman 2003.

[5] Davidson argues that "if we want to understand others, we must count them right in most matters" (Davidson 1974/1984, p. 197), but this only requires that beliefs be *mostly* true, it doesn't explain why *maximization* is required, that is, why, when faced with two interpretations according to which the subject's beliefs are mostly true, we should pick the interpretation that attributes to him or her *the most* true beliefs.

[6] For a more extensive discussion of Davidson's conception of the Principle of Charity, and how it relates to what this self-interpretation principle, see Jackman 2003.

[7] See Burge 1979. We are assuming for the sake of argument that the rest of his 'arthritis' beliefs were true of both *arthritis* and *tharthritis*, and that "treating as a mistake" is more substantial than simply giving up the practice of asserting the sentence. One could 'give up' one of the beliefs that one had arthritis in one's thigh because one wanted to bring one's usage into line with one's community without thinking that the original belief was false (just as I might "give up" my belief that someone moving from the United Kingdom

The Self-Interpretation Principle thus leads to a type of truth maximization and motivates it in terms of a form of self-knowledge that is underwritten by the fact that it is precisely the agent's point of view that we are trying to capture in interpretation. It leads to a type of maximization frequently associated with the Principle of Charity because it has always been assumed that what is to be maximized is not just the *total number* of true beliefs, but rather some *weighted* total of them.[8] Some beliefs will be more important to the speaker than others, and preserving the truth of these beliefs may have a higher priority than preserving the truth of multiple beliefs that are assigned less weight. It is precisely such weighing that would be revealed in the way the speakers revise their beliefs when conflicts become manifest,[9] and so assigning truth to the belief(s) that are preserved through such conflicts maximizes the (weighted) amount of truths the subject believes.

A holistic metasemantics that draws on the Self-Interpretation Principle can be shown to both leave meaning comparatively stable, and be compatible with semantic externalism,[10] and it will be argued here is that it also underwrites a type of *contextualism* that is given less emphasis in the literature than it could. In particular, if the function which maximizes the number of truths believed by the speaker works on weighted totals governed by something like the Self-Interpretation Principle, then we can account for some contextual variations in the semantic values of our words in terms of the fact that how much weight a particular belief has can vary from context to context.[11]

to the United States might "give up" his practice of referring to what American's call "cookies" as "biscuits" in order to be better understood in his new environment without thinking that any of his previous "that's a biscuit" utterances were mistaken).

[8] For more on the importance of such weighing in understanding anything like Charity as being remotely plausible, see Glüer 2011, Jackman 2003, Pagin 2006.

[9] Subject to some idealization relating to the fact that in practice speakers won't be aware of all of the incompatibility relations that exist among their various commitments, and so may in practice favor belief *A* over belief *B* simply because they were unaware that *A* was also incompatible with beliefs *B* and *C*.

[10] For an extended discussion of this, and whether the Self-Interpretation Principle should be understood as an explication or replacement of the 'Principle of Charity', see Jackman 2003.

[11] It is important to remember that the type of contextualism defended here is of a *metasemantic* sort. *Semantic* Contextualism treats the semantic value of a term as itself making reference to some feature of the context. Indexicals are paradigms of this ("I" means "the person making the utterance"), but accounts of, say, "tall" or "flat" that tie them to 'hidden parameters' (e.g. tall = 'tall for an f', where f is determined by the context) that are implicitly marked at the level of logical form do this as well. *Metasemantic* Contextualism, on the other hand, simply tells a story about how our terms get their semantic values that suggest that they may have different semantic values in different contexts (though the resulting semantic values themselves may make no reference to context). Consequently, one should thus not expect the sorts of syntactic markers associated with more familiar sorts of semantic

For example, we can see a type of context sensitivity by considering the following two sentences about a freckle-faced 7-year-old (Frank) who has just covered his school with graffiti.

(1) The principal should call Frank's mother.
(2) Frank probably gets those freckles from his mother.

Now if Frank is adopted, then (whether the speaker knows about the adoption or not) the extension of these two instances of "mother" will probably be different. (Assuming that we are acquainted with Frank but with neither of his 'mothers'.) We typically should contact the woman bringing Frank up if he is in trouble, but assume that the woman who contributed to his genetic make-up is responsible for his freckles. Somebody uttering the two sentences may thus refer to two different groups of people with the word "mother" not because they have added or lost any 'mother beliefs' but because the comparative importance of those beliefs changes from context to context.

The variance in the semantic value for "mother" can thus be understood as being produced by our various interests resulting in different 'mother beliefs' being more-or-less heavily weighed. When I am talking about Frank's disciplinary problems, my belief that Frank is being brought up by his mother will be weighed more heavily than my belief about his mother contributing to his genetic make-up. When I am talking about his freckles, the opposite will usually be the case. Different aspects of our 'mother prototype' are given greater weight in the two contexts.

An obvious alternative understanding of such cases would be to explain them in terms of ambiguity, that is, in terms of their being multiple words in one's lexicon that just happen to sound the same. Just as there are two distinct meanings for "bank," there would be multiple meanings for "mother." However, even if we were to admit that most of us had at least two separate "mother" entries in our mental lexicon,[12] there seem to

contextualism (for some discussion of these, see Stanley 2007) to be found if the contextualism involved is at the metasemantic level. Indeed, the contextualism defended here will be in many respects closer to the 'cheap' contextualism defended by Peter Ludlow (Ludlow 2008), and the view would suggest that the lexicon is 'dynamic' in much the way that he suggests (Ludlow 2008, 2014).

[12] And one might doubt whether two would be enough. Cases like the pair of sentences above could be created to show that "mother" would have to be ambiguous between the woman bringing up the child, the woman who contributes its genetic material, the woman who actually gives birth to the child, the woman who provides the egg, etc. We can certainly creat separate terms to explicitly distinguish birth mothers, biological mothers, traditional surrogate mothers, adopted mothers, stepmothers, and the rest, but it is hard to imagine that these are simply labels for separate items that existed in our mental lexicon as soon as we started talking about "mothers" at all.

be cases where this sort of response couldn't be appropriate (as when the two sentences were uttered by someone brought up in so sheltered an environment that they hadn't been aware that the two sorts of mothers ever came apart, and so had no reason to put two entries for the term in their mental lexicon). Further, even if one were to think that this sort of case is best handled by ambiguity, there are others where this seems harder to do.

For instance, consider a speaker who, when Venus rises at the beginning of the evening, will reliably point to it and say "there is Venus." He does the same in the morning, and never 'misapplies' the term to any other star or planet. The putative extension (those items that the term is *actually* applied to) of "Venus" in his language consists only of the planet Venus. However, while the putative extension of Venus in his language is much like it is in ours (assuming – falsely in my case – that we can identify Venus in the night sky), the general characterization he associates with the term is very different. In particular, while we believe that Venus is a lifeless planet, that we have sent satellites to gather information about it, and so on, this speaker believes that Venus is the goddess of love, that she is married to Vulcan, is the lover of Mars, responds to the prayers of her followers (of whom the speaker is one), and so on.

For such a speaker, there is a serious lack of fit between the general characterizations associated with "Venus" and what he actually applies the term to, and we could imagine his background to be shaped in three different fashions, each of which may lead us to interpret his use of "Venus" in different ways. If the speaker is a particularly devout worshiper, who was only interested in looking up at the night sky because he took himself to be viewing the distant goddess herself, then he may be much more willing to give up everything in the putative extension rather than his general beliefs. If he is informed that the distant star is a lifeless planet, he will be inclined, correctly, to think that he misapplied the term "Venus" every morning and evening when he looked at the sky.[13] On the other hand, if he were not at all devout, and his interest in 'Venus' was primarily as a guide to navigation, with the myths just providing interesting background about the light that happens to guide him, then he might give up all of his goddess-friendly general beliefs if he found them not to be true of

[13] There may be some temptation to treat utterances of his such as "Venus is looking especially bright tonight" as still being true (provided that the planet is especially bright on that night), but this may just be because (1) it is true in the sense that my utterance of "John looks hurt" might be considered true if I mistake Peter for John and Peter looks hurt, so that the speaker's reference for such utterances may be Peter or the planet, while the semantic reference remains John or the Greco-Roman goddess, or (2) one is implicitly moving on to a case more like the third type of Greek who shifts from one reference to the to the other depending upon what is most important to him in various contexts.

the celestial body he applied the term to, and so his use of "Venus," even in sentences like "Venus is a goddess," would refer exclusively to the planet. Finally (and this may have put him in the largest group at some point), he may be someone in between, whose goddess-friendly beliefs are more important in some contexts, and whose navigation-friendly beliefs are more important in others. In this final case, it may be that what he refers to by the term switches form context to context, sometimes picking out a planet, sometimes a fictional goddess, and so on.[14]

It should hopefully be clear at this point how this sort of metasemantic context-shifting is intended to work,[15] and it will be argued below that we see a similar kind of context-shifting in a number of familiar philosophical thought experiments, and a lack of recognition of the 'shiftiness' of these cases can lead to the intractability of the disputes about what to say about them.

Metasemantic contextualism, ambiguity, and externalist intuitions

The literature on semantic externalism is often driven by intuitions about particular cases (finding "water" on Twin Earth, Bert's complaining to his doctor about his "arthritis", etc.), but philosophers have never

[14] In much the same way, who, if anyone, we refer to by "Moses" may depend upon how heavily our beliefs are weighed. Someone who is interested in the history of the Middle East may weigh heavily the belief that Moses led the Israelites out of Egypt, but not put much weight on beliefs relating to the miracles Moses purportedly performed (even if he does believe that they were, in fact, performed). On the other hand, someone who is only interested in the miraculous aspects of the story might weigh the miracles the most heavily, and if there turns out to be no one who took the tablets containing the Ten Commandments from God and parted the Red Sea, then his term "Moses" would not refer to anyone. Finally, someone may be both of these, interested sometimes in history, and sometimes in miracles, and the reference of the name may shift for him accordingly. (Whether something like this is what Wittgenstein was getting at when he claimed that "If one says 'Moses did not exist', this may mean various things" (Wittgenstein 1953, §79), is something that I'll leave to the reader to decide.)

[15] One may note similarities here to the view defended in Bilgrami (1992), which also allows for a type of metasemantic context sensitivity. However, Bilgrami distinguishes 'aggregate contents', which capture *all* of the beliefs an agent associates with a term, from the more psychologically real 'local contents' which are made up a contextually salient *subset* of the beliefs in the aggregate contents, and thus vary from context to context. Context determines not how much weight a particular belief has in determining what we are referring to, but rather whether the beliefs in question are part of a particular 'local' content *at all*. The ultimate goal of Bilgrami's theory is precisely not to explain how our beliefs determine what we refer to, but rather to understand meaning in terms of belief rather than reference (see especially Bilgrami 1992, pp. 134–35) while preserving the intuition that meanings can be shared (because, even if the 'aggregate' contents are never shared, the local subsets contextually derived from them can be; for a discussion of this, see Jackman 2014).

reached complete agreement about such intuitions, and it has recently been sugested that nonphilosophical intuitions about such cases are generally much more variable than most philosophers had originally assumed.[16] That said, if the kind of contextualism outlined above is right, then the way in which we describe our cases will be very important, and an intuition about what a word refers to in one case need not entail that it refers the same way in others. Consequently, what are often described as "conflicting intuitions" may just be intuitions about different cases that conflict when treated as context free judgments about what our terms refer to, but are perfectly compatible with each other when their context is made more explicit.

For instance, let us return to Tyler Burge's discussion of Bert and his idiosyncratic use of "arthritis." Bert uses "arthritis" much as the rest of us do but, notoriously, he also applies "arthritis" to the pains in his thigh.[17] Burge treats Bert here as still referring to *arthritis* by "arthritis," while writers such as Davidson and Bilgrami argue that Bert should be understood as referring instead to *tharthritis* (once again, a condition which includes both arthritis and similarly painful ailments in the limbs) by "arthritis."[18] This lack of consensus about what to say about Bert may reflect the fact that what Bert means by "arthritis" varies from context to context. When Bert goes to the doctor and complains "my arthritis has spread to my thigh," it may be correct to take him to be referring to *arthritis* by "arthritis." On the other hand, when he is sitting around with his brother and complains "my arthritis is too bad for me to mow the lawn today," it may be equally correct to treat him as referring to *tharthritis*.

The sort of metasemantic contextualism outlined above would explain why the extension of Bert's term "arthritis" might shift in just this way. In addition to a large set of beliefs which would be true of both *arthritis* and *tharthritis*, Bert has one set of beliefs (such as "I have arthritis in my thigh" and "My arthritis kept me from cleaning out the garage last week") which would be true only of *tharthritis*, and another set of beliefs (such as "Doctors have studied how to treat arthritis" and "The man from the insurance company said that people with arthritis should go see a doctor

[16] See, for instance, Weinberg, Nichols, and Stich 2001; and Machery, Mallon, Nichols, and Stich 2004.

[17] See Burge 1979.

[18] Bilgrami 1992; Davidson 1993, 1994. Davidson seems to have reconsidered his views on the example, though only if Bert "intended his hearers to take the word 'arthritis' as referring, not to what he thought it referred to, but to what it referred to when used by experts" (Davidson 2003, p. 698). Since this intention is also supposed to be "explicit," it's not clear how close to Burge's view he actually comes, because Davidson's requirement that this intention be "explicit" seems to suggest a metasemantic understanding of the example that Burge explicitly rejects (see Burge 1979, pp. 96–97).

about it") which would be true only of *arthritis*. When he is complaining to his brother, the former set of beliefs will be given greater weight than the latter (and so he will refer to *tharthritis*), while when he is consulting his doctor, the latter set of beliefs will be given greater weight (and so he will refer to *arthritis*). What Bert means by "arthritis" shifts from context to context, and such context sensitivity may be characteristic of cases where someone's idiosyncratic usage of a word can be understood in terms of either an idiosyncratic belief or an idiosyncratic semantic value.

It would, then, not be a coincidence that Burge tends to focus on examples like visits to the doctor's office, while Davidson and Bilgrami present cases where our interests are less tied to professional standards. Each presents Bert using of "arthritis" in a context that is friendly to their views, and then generalizes from that usage to what Bert means by "arthritis" in a more context-free way.[19] Furthermore, conflicting intuitions about what Bert means may also reflect the fact that, if the comparative entrenchment of one's beliefs is part of what determines what one means, then cases like Bert's use of "arthritis" are underdescribed in a way that allows different readers of the stories to project their own sense of which beliefs should be most important on to the characters involved in them.[20] So, for instance, if we are personally disinclined to defer to expert usage even in situations where that expertise is being relied on, we may simply project those weightings on to Bert, and so, as Davidson often does, treat him as meaning *tharthritis* by "arthritis" even when he is in the doctor's office.

Much the same sort of move could be made with respect to the sometimes conflicting intutions we have about terms like "water."[21] We have

[19] Though the fact that Bert would come out as meaning *arthritis* even in *some* cases would be enough to establish anti-individualism, so this sort of contextualism would side more with Burge than Davidson on the larger issue at hand. Indeed, one could argue that even in those contexts where we are more inclined to say that Bert means *tharthritis*, it still may fail to be the case that that those contents are individuated individualistically. If Bert wants to get out of mowing the lawn because of the pain in his thigh, but lived in a society that used "arthritis" to pick out, say, everything philosophers associate with "tharthritis" *except for* pain at the base of the spine (call it 'barthritis'), I'd expect that Bert would be correctly taken to mean *barthritis* rather than *tharthritis*, even if he happened to believe that one could get 'arthritis' at the base of one's spine. Consequently, even when an idiosyncratic concept is attributed to Bert, it doesn't follow that the concept he has is independent of how the relevant word is used in his society. (For a related discussion of why nonindividualism holds for speakers who have mastered the concepts in question, see Burge 1979, pp. 84–85.)

[20] See, once again, Jackman 2009.

[21] Though I think that the intuitions are less often in conflict for philosophers like myself who have had the importance of "water" being a natural kind term drilled into us since the beginning of our undergraduate education. The variance would be even more pronounced when one considers the use of the term before the discovery of modern chemistry.

many commitments that focus on water as a functional kind, and others that focus on it as a natural kind, and while we often presume these kinds to be co-extensive, the two would come into conflict if we were ever in a scenario like Putnam's Twin Earth case, in which we were confronted with a substance that had all of water's perceptual/functional properties (its appearance, its taste, its ability to sustain life, etc.) while having a different chemical structure. Which commitments will be weighted most heavily, and so whether "water" picks out a natural or functional kind, could vary from person to person, and with certain individuals, from context to context. Indeed, this seems to be compatible with Putnam's own analysis of the extension of "water":

> We can understand the relation *sameL* (same liquid as) as a cross-world relation by understanding it so that a liquid in *W1* [World 1] which has the same **important** physical properties (in *W1*) that a liquid in *W2* possesses (in *W2*) bears the *sameL* to the latter liquid . . . an entity *x*, in an arbitrary possible world, is *water* if and only if it bears the relation *sameL* (construed as a cross-world relation) to the stuff *we* call 'water' in the *actual* world. (*Putnam 1975, p. 232; boldface mine*)

Putnam claims that the 'hidden structures' determine the reference of natural kind terms not because only such hidden structures could serve in the same-kind relation, but rather because "normally the 'important' properties of a liquid or a solid, etc., are the one's that are structurally important." However, importance is, as Putnam himself goes on to stress, "an interest relative notion" (ibid., p. 239), and in some contexts the more functional properties associated with "water" may be more important to us than its microphysical ones.[22] In such contexts, it may be that the term is best seen as picking out the functional kind. Furthermore, when conversing with others, or when reading a philosopher's paper on the subject, we might not intially have a strong preference between the commitments tying to each of the two readings, and thus be willing to favor one set over another simply to 'accommodate' our conversational partners.[23]

This 'pluralistic' attitude toward the apparently conflicting intuitions outlined above bears some similarities to a view currently being proposed by Shaun Nichols, Ángel Pinillos, and Ron Mallon, who also

[22] This is certainly downplayed in Putnam 1975, but a sensitivity to how we can still feel the pull of the nonnatural kind reading in such cases is more evident in Putnam 1962. (For a discussion of Putnam's 'hardening of heart' on this issue, see Unger 1982, p. 165; Unger 1984 p. 124).

[23] This seems to be the understanding of such cases in Lewis 1999, pp. 313–14.

stress how our judgments about the standard externalist thought experiments are much more variable than many philosophers assumed.[24] However, they suggest that the variance between intuitions about what to say about various thought experiments in the philosophy of language should be explained by something closer to full-scale ambiguity. As they put it:

> We will argue that natural kind terms are ambiguous. In some cases, the reference of a token is fixed by a causal-historical convention; in other cases, the reference of a token of the same type is fixed by a descriptivist convention. We call this an ambiguity theory because the idea is that there are two conventions that determine the reference of natural kind terms . . . (*Nichols, Pinillos and Mallon 2014, p. 7*)

While the authors take the ambiguity in question to be less extreme than the sort we see with, say, "bank,"[25] where the two semantic values are unrelated, it still seems that the ambiguity they have in mind is considerably more substantial than the sort of context sensitivity defended here. In particular, they suggests that there are two completely different ways ("two conventions") in which the referents of our terms can be determined. As they put it, their claim that natural kind terms are systematically ambiguous between descriptive and nondescriptive conventions runs against a "critical constraint on theory building in the philosophy of language," namely that "only one theory of reference will apply to a class of terms."[26]

Now the contextualist account above is not only one that respects this "critical constraint" by proposing a "univocal theory of reference"[27] (the same function from use to meaning would determine what, say, "Arthritis" refers to in both cases, its just that the weighting of the inputs to the function changes), but it is also one that allows a good deal more flexibility than one that took reference to simply be ambiguous between causal and descriptive notions. In particular, taking our terms to be ambiguous between the 'causal' and 'descriptive' meanings doesn't help account for contextual shifts in which both meanings seem to fall within the range of the causal. In short, the contextualist account can, in principle, explain all the cases appealed to by the ambiguity theory, and others

[24] Nichols, Pinillos and Mallon (2014). The authors also make the suggestion that philosophers tend to cherry pick their examples to favor those which produce intuitions favoring their preferred theory (see ibid., p. 10).

[25] Ibid., p. 26. Indeed, they also appeal to Lewis 1999 (see note 21), but they seem to treat the "accommodation" involved as deciding which of the two referential conventions to follow, rather than involving which of the commitments associated with a word should be given more stress (Nichols, Pinillos and Mallon 2014, p. 24).

[26] Nichols, Pinillos and Mallon 2014, pp. 22–23.

[27] Ibid., p. 19.

besides, so it is unclear what advantage positing the more substantial ambiguity to explain for such cases would bring.[28]

'Switching' cases

Cases of contextual variation that can't be explained in terms of a simple ambiguity between 'causal' and 'descriptive' conventions arise when the source of information associated with a word that the subject is causally connected to 'switches'. We will focus on three such cases here.

Donnellan and Evans on proper names

An example of this sort can be seen in the contrasting views of Evans and Donnellan about how to understand the reference of a proper name when the source of the information associated with it changes over time. Donnellan allows that the reference of such names can vary from context to context, while Evans favors a more invariant account. I'm largely sympathetic with the general shape of Evans's theory, but he mischaracterizes Donnellan's position in a way that keeps him from seeing how his own view could be developed. In particular, Evans takes Donnellan to argue for something like what I'll call "source-dependent contextualism," which is roughly, the view that a token of a name in a claim refers to whichever potential bearer was the source of the information that is the topic of that claim.

For the source-dependent contextualist, if our beliefs about "Napoleon" actually came from two men, one of whom ("Alpha") was the source of all of our pre-1814 information associated with "Napoleon," and another ("Beta") of whom took the original's place and was the source of the rest of our information, then claims about "Napoleon's" early life or the 1812 invasion of Russia would refer to Alpha, while claims about "Napoleon's" defeat at Waterloo or his eventual death on the island of Saint Helena would refer to Beta. By contrast, Evans argues as follows:

> I think that we can say that *in general* a speaker intends to refer to the item that is the dominant source of his associated body of information. It is important to see

[28] The contextualist account also explains the fact (discussed ibid., p. 20) that many questions elicit a type of 'neutral' reaction rather than showing a response that clearly favors either the causal or descriptive reading. The ambiguity view might suggest that we could toggle between these two distinct sets of conventions, but the contextualist view suggests that we may face cases where the incompatible beliefs are equally weighted (or in which we have no idea what weight they have) and so have no firm idea of what is being referred to.

> that this will not change from occasion to occasion depending upon the subject matter. Some have proposed [Donnellan 1970 is cited at this point] that if in the case [above] the historian says "Napoleon fought skillfully at Waterloo" it is the imposter Beta who is the intended referent, while if he had said in the next breath '. . . unlike his performance in the Senate' it would be Alpha. This seems a mistake; not only was what the man said false, what he intended to say was false too, as he would be the first to agree; it wasn't Napoleon who fought skillfully at Waterloo.[29]

Evans doesn't think that the reference of "Napoleon" shifts from context to context, but while he is correct to claim that Donnellan does allow for contextual variation, he doesn't present a charitable, or accurate, reading of the sort of variation that Donnellan makes room for. Indeed, Donnellan's claims about what we would be referring to in various situations seem to be a better fit with the type of contextualism defended here, than they do with the position Evan's attributes to him.[30]

We can see this by reconsidering the primary example from the paper of Donnellan's that Evans cites. In "Proper Names and Identifying Descriptions," Donnellan asks us to imagine that a student attending a party "meets a man he takes to be the famous philosopher, J. L. Aston-Martin."[31] The student has previously read a number of Aston-Martin's papers and spends an hour or so speaking with the man (but not about philosophy) over the course of the evening. Donnellan claims that if the student goes back to his seminar the next day and tells everyone "Last night I met J. L. Aston Martin and talked to him for almost an hour," he would have said something false, since the name refers to the philosopher, and not the man a the party.[32] On the other hand, if he is telling some other friends about the party and the amusing things that happened, such as when "Robinson tripped over Aston-Martin's feet and fell flat on his face," the name would refer to the person at the party, not the famous philosopher.[33]

First of all, one should note that on the account attributed to Donnellan by Evans, *both* of these utterances of "Aston-Martin" would have referred to the man at the party (while it would refer to the philosopher when the student said things like "J. L. Aston-Martin wrote 'Other bodies'"), while in Donnellan's own discussion the referent of the name *isn't* identified with the source of the particular belief that is forgrounded in the utterance. Instead, Donnellan tentatively explains the difference in who the name refers to in the two cases as follows:

[29] Evans 1973, pp. 16–17.

[30] For a more sympathetic reading of what is motivating Donnellan here that meshes with much of this, see Stalnaker 2008 (especially pp. 124–25).

[31] Donnellan 1970, p. 68. [32] Ibid., pp. 68–69. [33] Ibid., p. 69.

> Perhaps the difference lies in the fact that the initial utterance of the speaker's remark would only have a point if he was referring to the famous philosopher, while in the later utterances it is more natural to take him to be referring to the man at the party, since what happened there is the whole point.[34]

The referent of a name may switch from context to context because different sources of information can, in Evans's terms, 'dominate' the total "body of information" associated with the name, and this domination may vary depending on what the point of a particular utterance is. Chances are that piece of information will bear greater weight than usual if it is the topic of the utterance, but this alone would not guarantee that other information associated with a name might carry still greater weight. If there are two sources of information associated with a name, one of which accounts for the great majority of the beliefs involved, it is quite likely that one will be 'dominant' even in cases where the sentence topic is something tied to the other (as in Evan's "Napoleon" case). By contrast, if the amount of information is more evenly split, then the 'dominant' source may often vary from context to context. Still, even if things are generally weighted heavily one way, what becomes dominant can switch. For instance, in a note to the passage quoted above, Donnellan considers a case where the student becomes good friends with the "Aston-Martin" he meets at the party, and stays so for many years without ever discovering that he is mistaken about his friend's philosophical production. In such cases, most of the speaker's "Aston-Martin" utterances will refer to his new friend, since information from the party-going Aston-Martin would typically dominate the rest, but if he claimed to know J. L. Aston-Martin "in circumstances where it is clear that the point of the remark has to do with claiming to know a famous man," Donnellan still thinks that "we would suppose him to have referred to Aston-Martin, the famous philosopher, and not to [the] man he met at the party, who later is one of his close acquaintances."[35]

Such shifts could not, of course, be explained in terms of the meaning of proper names being 'ambiguous' between a 'causal' and a 'descriptive' reading, since both interpretations of the name are equally causal. The use of "Aston-Martin" to pick out the famous philosopher rather than the party-goer isn't purely "descriptive," as should be clear when we consider the possibility that the famous Aston-Martin fabricated much of his back story and achieved his fame by taking credit for the work of one is his more retiring colleaugues (and so never wrote "Other Bodies" or any of the other papers attributed to him).[36] There might, of course, be contexts in which the term is used in ways that make writing the papers more essential

[34] Ibid. [35] Ibid., p. 79. [36] See the discussion of "Gödel" in Kripke 1972/1980.

(as when a student is writing his dissertation on "Other Bodies" but cares little about the biography of the author), but that wouldn't be one of the contexts that Donnellan describes, where the "famous philosopher" and the "party-goer" are the two subjects.

Memory content

The account sketched above also applies usefully to the problem of how to ascribe thought (and memory) content in cases such as the following:[37] John is, at the age of 12, transported from Earth to Twin Earth (which is, once again, just like Earth but whose 'water' is made up of XYZ (*twater*) rather than H_2O) and he lives there for another forty years without being aware of the switch.[38] Most externalists agree that at age 12, John's use of "water" refers to H_2O but at some time over the next few years his term comes to pick out (or at least typically pick out) XYZ, so that when he asks for a glass of "water" he is talking about *twater*, and no longer talking about water. There is, however, less consensus about whether John has (1) simply acquired a *second* 'water' concept, so that he is able to have thoughts about both *water* and *twater* (the pluralistic, or "conceptual addition," view),[39] or (2) had his original *water* concept *replaced* by a *twater* concept, so that he is now unable to have any *water* thoughts (the monistic, or "conceptual replacement," view).[40]

The difference between the pluralistic and monistic views manifests itself when we try to interpret claims/thoughts of John's (at age 52) such as "I remember swimming in Lake Ontario when I was 11 and thinking 'this water was freezing!'" Defenders of the monistic view typically claim that in such attributions "water" picks out *twater* and that John has simply lost the ability to remember what he thought before. Defenders of the pluralistic view, on the other hand, may treat this as one of the cases where John was able to apply his original *water* concept, so that the recollection turns out to be a true one. Monists view even memory content to be determined by one's current environment, while pluralists typically treat

[37] Such cases of 'slow switching' became familiar through Burge 1988 and Boghossian 1989. For a summary of some of the discussion of such cases, see Parent 2013.

[38] This is the most familiar version of the case, but the problem can be generated around less far-fetched examples such as someone making a permanent move from Great Britain to the United States without realizing that the term "Robin" picks out different birds in the two dialects of English, and unable to tell the two sorts of bird apart.

[39] See, for instance, Boghossian 1989, 1992a, 2011; Burge 1988; Gibbons 1996; Heal 1988.

[40] See, for instance, Bernecker 2010; Brueckner 1997; Ludlow 1995, 1996, and 1999; Tye 1998.

memory content as reflecting the environment in which the thought remembered originally occurred.

The view defended here is clearly more pluralistic than monistic. However, it is in a position to allow that the monist may often be right about what we should say about particular memories. The contextualist view allows that there are a variety of 'water' concepts available to the speaker in the switching cases, and which one, say, John applies will depend upon his interests at the time. At 52, John typically refers to *twater* by "water" since most of his 'water' beliefs are tied to his contact with XYZ. However, if the purpose of his recollection were simply to reflect on his youth, most of these later water beliefs may not be relevant, while the belief that he did, in fact, experience the freezing water would be very heavily weighed. In such a case, he would plausibly be seen as thinking about *water* rather than *twater*.

On the other hand, if the recollection comes up in the context of debating whether a spot he and his family are about to visit (and which he thinks he once swam in) will be pleasant to swim in, then non *water*-friendly beliefs of his such as "there is a Great Lake full of water to the north of me that I'm thinking of vacationing at," "my children always complain when they have to swim in cold water," and so on, will be heavily weighed, and the suggestion that he refers to *twater* by the term[41] (and thus misremembers what he originally thought) will seem much more plausible.

Whether a speaker's use of "water" in a memory claim refers to H_2O or XYZ will depend at least partially on his context, and not simply on the source of the particular memory claim. In those contexts where his commitments associated with his original environment have more weight, it will refer to *water* while in those where those relating to his new environment have more weight, it will refer to *twater*. Of course, as the speaker spends more time on Twin Earth, the number and strength of his *twater* commitments is bound to grow, but the *water* commitments don't simply disappear, and in some (increasingly rare) contexts they may be important enough to outweigh the more recent *twater* commitments.[42]

There may also, of course, be contexts where "water" referred to a more 'functional' kind that picked out both H_20 and XYZ. In such

[41] Or perhaps a functional kind that picks out both H_2O and XYZ (see below).

[42] So, this would follow Heal in suggesting that the switch to the new contents would never be 'complete'. The original substance would be what was referred to not just because those instances of a kind that "would be the first to come to mind" in the context (Heal 1998, p. 105), but also because they would be the ones the speaker took to be most important for the point he was trying to make if questioned.

contexts, the commitments to applications on both planets would each be more heavily weighted by the speaker than any commitments to his using the term in just the way his peers do,[43] or to the term picking out a natural kind (as, perhaps, when the speaker insists that "Ever since I was 7, I've felt 'off' all day if I don't start the morning by splashing cold water on my face").[44]

Pluralistic accounts of switching cases are sometimes criticized for not suggesting any sort of "mechanism" that would determine which contents occurred in which contexts.[45] Monistic accounts have no such requirement, but the sort of contextualist account suggested above has no trouble with this extra explanatory burden. The Self-Interpretation Principle provides an explanation of why and when the contents switch, and it does so in a way that keeps self-knowledge centrally located in our account of such contents.[46] Of course what I described earlier as "source-dependent contextualism," the view that would simply tie the content of "water" in a memory claim to whatever substance was the source of that particular 'water' memory, also provides a mechanism for determining which contents appear in which contexts,[47] but that view not only makes unintuitive claims about just what the content of our memories must be in certain contexts (see the discussion of Evans and Donnellan above), but, as we shall see in the following section, it also leads to some serious tensions with some natural assumptions about self-knowledge.

[43] Especially since, given that the functional kind includes all of the instances of the natural kind they are talking about, his meaning something different than they do doesn't necessarily prevent them from being legitimate sources of testimonial knowledge for him.

[44] Once again, given the multiplicity of possibly readings, these switches don't lend themselves to any sort of easy explanation in terms of "water" being ambiguous between a 'descriptive' and a 'causal' reading.

[45] See, for instance, Bernecker 2010, p. 192.

[46] The predictions of such an account would be largely in line with the other account that I know of to give a plausible account of how the plurality of contents would be sorted, namely, Goldberg's (2005) account of what he calls "the defeat of the Current Face Value Presumption." Roughly put, Goldberg's view is that a speaker's term should be interpreted in accordance with the meanings of his current community unless (1) the speaker's intentions and justificatory dispositions regarding that word support a different interpretation, and (2) the speaker would, on becoming aware of this conflict with what his current society means by the term, disavow any intention to be using the word with that meaning, and correct his other beliefs accordingly. However, while Goldberg's account would make roughly the same predictions about cases (1) and (2) essentially amount to the speaker having commitments tying his use to the original context that are more heavily entrenched than those tying his use to his current context, it doesn't quite answer Bernecker's demand for a mechanism explaining why various contents appear in various contexts. That is to say, it describes how the semantic values may be distributed without giving any sort of metasemantic explanation of the distribution.

[47] Though it is less clear how the account can extend beyond direct memory reports, or reports that draw on multiple memories with multiple sources.

Externalism and inference

The understanding of the switching cases available to the metasemantic contextualist also has the advantage of allowing one to put to rest Boghossian's worry that "externalism is inconsistent with very important aspects of our intuitive conception of the mind – namely, with the a priority of our logical abilities."[48] Given that the breakdown in the priority of our logical abilities that Boghossian's argument focuses on seems to turn on our apparent inability to recognize that two concepts in our occurrent thoughts have different contents, the breakdown in our logical abilities seems tied to a breakdown in some more intuitive sense of the 'transparency' of our own thoughts, and hence suggests a serious tension between semantic externalism and our intuitive conception of self-knowledge.[49]

Boghossian argues for his conclusion by considering the following variant of the now familiar switching case: Peter is an opera fan and inhabitant of Earth. While vacationing in New Zealand he encounters Luciano Pavarotti swimming in Lake Taupo and, much to his delight, has an extended conversation with the famous tenor. Some time later, and without his knowledge, Peter is switched to Twin Earth and over a number of years most of his terms come to take on the semantic values standardly associated with Twin English, so that when he eventually sees Pavarotti's twin perform and subsequently claims that he saw "Pavarotti" sing, he is talking about *Twin* Pavarotti. However, it seems as if he can still have memory-based thoughts about the Pavarotti on Earth, and when he reminisces about the time that he saw 'Pavarotti' swimming in Lake Taupo, he seems to be thinking of our Pavarotti, not his twin. Boghossian claims that this ability to access both contents could, however, result in Peter's engaging in reasoning like the following:

(1) Pavarotti once swam in Lake Taupo.
(2) The singer I heard yesterday is Pavarotti.
(3) Therefore, the singer I heard yesterday once swam in Lake Taupo.

[48] Boghossian 1992a, p. 17. Ludlow (1999) appeals to similar cases to undermine the pluralistic or 'conceptual addition' interpretation of the switching cases, since this particular problem doesn't arise for those who endorse the monistic or 'conceptual replacement' interpretation of such cases.

[49] Boghossian later cashes out the relevant sense of 'transparency' as: "If two of a thinker's token thoughts possess the same content, then the thinker must be able to know a priori that they do; and (b) If two of a thinker's token thoughts possess distinct contents, then the thinker must be able to know a priori that they do" (Boghossian 2011, p. 457).

According to Boghossian, "Pavarotti" refers to Earth's Pavarotti in the initial premise, but to Twin Pavarotti in the second, so that "True premises conspire, through a fallacy of equivocation that Peter is in principle not able to notice, to produce a false conclusion."[50] What would seem to Peter to be a perfectly valid inference would thus turn out not to be and so "the thesis of a priority of logical abilities is . . . inconsistent with externalist assumptions."[51] This is a surprising result, and it (along with its suggestion that the identity of our concepts is not 'transparent' to us) is sure to make externalism look less appealing.

However, there is reason to be suspicious of Boghossian's claim that the references of our terms could switch mid argument in such a fashion. In particular, Boghossian's argument seems to rely on something like the 'source-dependent contextualism' mentioned earlier,[52] and thus he feels free to ignore the fact that arguments and inferences (especially those inferences meant to be introspectively surveyable) take place against the background of a single context.

In particular, just because Peter might typically utter (1) in contexts where the occurrence of "Pavarotti" refers to Pavarotti, while he typically utters (2) in contexts where the occurrence of "Pavarotti" refers to Pavarotti's twin, it need not follow that the term can support multiple semantic values in the context of an argument that involves both (1) and (2). Indeed, the contextualist account outlined above would seem to rule out the sorts of introspectively undetectable equivocation described by Boghossian.[53]

Peter has, after all, a set of 'Pavarotti commitments' that are tied to two men, and which of the two Pavarottis he refers to will depend on which subset of his commitments carries the most weight in a given context. It seems likely that a sentence like (1) would typically be uttered in the context of Peter's reminiscing about his encounter with the great tenor, and the commitments tying the term to Pavarotti will in such cases be more heavily weighted than those attached to his twin. (Of course, this

[50] Boghossian 1992a, p. 22. (We can assume that Pavarotti's twin never swam in either Lake Taupo or its twin.)

[51] Ibid.

[52] As he puts it, "memory-beliefs involving 'Pavarotti' or 'water' or whatever, originating in Earthly experiences, are about Pavarotti, water, and so on. And, correlatively, that memory-beliefs involving those words originating in Twearthly experiences will be about Twin Pavarotti, twin water, and so on" (Boghossian 1992b, p.42.)

[53] Stalnaker (2008) also presents an account of these slow-switching cases that doesn't rely on any type of source-dependent contextualism, but his view seems to give priority to the *attributer's* context (see especially p. 131), in a way that this view does not, and thus may be open to criticism (see, for instance, Boghossian 2011, p. 465) for taking these intentional facts dependent upon other people's intentional facts, which would, in turn, face just the same need for interpretation.

need not *always* be the case, and (1) could be uttered in contexts where the commitments associated with Pavarotti's twin carried more weight.) By contrast, a sentence like (2) would be more likely uttered in contexts where the commitments tied to Pavarotti's twin are more entrenched, and in that case "Pavarotti" would refer to Pavarotti's twin. (Though, once again, there could be contexts where the occurrence of the name in (2) might refer to Pavarotti as well.)

Still, while there can be contexts where "Pavarotti" refers to Pavarotti, and other contexts where the name picks out Pavarotti's twin, these are clearly *different* contexts, and while isolated instances of either (1) and (2) can each be true in some context, there is no context in which they are *both* true. However, when (1) and (2) are incorporated into a *single argument*, they have to be interpreted in terms of *single* contexts, and so the term will need to be assigned a single semantic value in both (1) and (2).[54] Consequently, the sorts of hidden equivocation (and corresponding failures of semantic self-knowledge) that Boghossian treats the externalist as committed to allowing will not arise.

The view presented above allows both that (1) the thinker has access to two "Pavarotti" contents, and (2) those two contents will never produce an equivocation in the course of a single argument.[55] Boghossian is skeptical about the possibility of such a combination, but such skepticism seems driven by the assumption that such a combination could only be explained by a view (which he attributes to Burge and Schiffer)[56] in which the terms in an argument are quasi-anaphorically connected so that the first occurrence of a term (whose reference is determined along source-dependent contextualist lines) itself determines what the rest of the occurrences of that term refer to.[57] On such accounts (as Boghossian notes), the ordering of the premises would determine who the argument is about, so that

[54] See Schiffer 1992, p. 35 for a similar point and the claim that premises 1 and 2 can be compressed into something like "Pavarotti once swam in Lake Taupo, and is the singer I heard yesterday" where there is only one occurrence of "Pavarotti."

[55] For some others who deny that "Pavarotti" would have two distinct references in this case, see Burge 1998, Goldberg 2007a, and Schiffer 1992.

[56] Burge 1998 and Schiffer 1992. Note, however, that while Schiffer insists that "Pavarotti" in the second premise be interpreted in the same way as it is in the first (Schiffer 1992, p. 33), his presentation seems to leave room for the possibility that this might be a consequence of the fact that whatever determines the content of one will determine the content of the other in the context of an argument. Consequently, he doesn't seem to explicitly commit himself to the *order* of the premises determining how they are interpreted.

[57] Boghossian 2011, p. 459. For a similar argument, see Brown 2004a, pp. 177–78, and see Bernecker 2010, p. 191, and Goldberg 2007a, p. 181 for other discussions of this sort of 'anaphoric' explanation of univocality within arguments.

(1) Pavarotti once swam in Lake Taupo.
(2) The singer I heard yesterday is Pavarotti.
(3) Therefore, the singer I heard yesterday once swam in Lake Taupo.

would be a valid argument about the Pavarotti on Earth (with a false second premise), while

(1) The singer I heard yesterday is Pavarotti.
(2) Pavarotti once swam in Lake Taupo.
(3) Therefore, the singer I heard yesterday once swam in Lake Taupo.

would be a valid argument about the Pavarotti on Twin Earth (with a false second premise). That the ordering of premises should have such an effect is an admittedly unappealing conclusion, but while that presents a problem for those who respond to Boghossian's argument with such an appeal to quasi-anaphoric dependence, it should be clear that the sort of metasemantic account suggested here has no such commitment. There is no requirement that the sentence in the first premise be weighted more (or less) heavily than the sentence in the second premise, so while all of the occurrences of "Pavarotti" in the argument must be assigned the same value, that value is independent of the premises' ordering.

Conclusion

The metasemantic contextualist who relies on something like the Self-Interpretation Principle can, then, argue that a number of important debates surrounding semantic externalism arise from the fact that both sides mistakenly assume that there is a single answer to questions like "What is the semantic value of 'arthritis' in Bert's language?" or "What is the content of John's 'water' memories?", when the answers to such questions are, in fact, context-dependent. Further, while the view is motivated by a certain conception of self-knowledge as a type of 'authority' (that is, speakers' own judgments determine what they mean when their commitments conflict), it is able, once in place, to help answer the threat to self-knowledge as 'transparency' that semantic externalism seemed to present.