Reference Magnetism Does Exist: A Reply to Warren

Peter Finocchiaro[1]

Abstract

In a recent paper, Jared Warren argues that reference magnetism doesn't exist. His argument rests on the claim that reference magnetism is arbitrarily weak. In this paper, I show why Warren's argument fails. Reference magnetism is arbitrarily weak in some contexts. But Warren seems to assume that the force of reference magnetism cannot change across contexts. In fact, though, the force of reference magnetism is something that we change. Thus, reference magnetism does exist – when we make it exist.

Keywords: Metaontology; metasemantics; reference magnetism; charity; context-sensitivity; semantic externalism

Introduction

If reference magnetism exists, then, sometimes, our words mean one thing rather than another in virtue of the fact that the first thing is more metaphysically natural than the second thing. The word 'electron', for instance, refers to all electrons rather than only the electrons that are outside the event horizon of black holes because *electron* is a natural kind and *electron outside the event horizon of black holes* is not.

In a recent paper, Jared Warren has argued that reference magnetism does not exist. If Warren's argument succeeds, then many metaphysicians are in trouble insofar as we rely on reference magnetism to establish the substantivity of what we do.

Take ontological disputes as a case in point. Some metaphysicians claim that tables exist and others claim that they do not. Someone like Eli Hirsch would say that they are both right in their own way (see, *inter alia*, Hirsch (2009): 239–241). The furniture realist's use of the word 'exist' should be interpreted expansively so as to include tables, the furniture nihilist's use of the word 'exist' should be interpreted restrictedly so as to exclude tables, and the two metaphysicians

[1] **This is a draft.** Try not to needlessly cite or criticize unless your goal is to professionally embarrass me. (If your goal is to professionally embarrass me, then let me know and I'll give you some more effective strategies.)

should acknowledge this interpretive difference. Thus, the two metaphysicians should acknowledge that their dispute turned out to be merely verbal. Someone like Theodore Sider would respond to Hirsch's challenge by insisting on a shared meaning of 'exist' (see, *inter alia*, Sider (2011): 54–63). When the two metaphysicians use the word 'exist', both of them should be interpreted such that 'exist' means whatever available meaning is most natural. The metaphysicians may not know what that meaning is. They may even claim that it is something other than what it actually is. Nevertheless, reference magnetism makes 'exist' mean the natural thing rather than the thing they think it means.

This response to Hirsch's challenge becomes unavailable if Warren's argument succeeds. Thankfully, Warren's argument does not succeed. That being said, Warren's argument does succeed as a reminder about some aspects of reference magnetism that we tend to neglect. In particular, Warren's argument reminds us that metasemantics isn't immutable: the force of reference magnetism can change, and we metaphysicians need to be more mindful of how it can change.

I begin by summarizing Warren's argument in Section 1. Then, in Section 2, I present some cases that complicate Warren's overall conclusion. Finally, in Section 3, I sketch a theory of reference magnetism in order to respond to Warren's argument. My sketch is far from being a complete theory and much more remains to be said. Yet I think the sketch is enough to show why Warren's argument fails.

1 - Warren's Argument

Warren's argument for the claim that reference magnetism doesn't exist can be helpfully divided into two parts. In the first part of the argument, Warren presents a wide range of cases in which it appears that reference magnetism is arbitrarily weak – that is to say, reference magnetism never justifies interpreting an utterance contrary to how it was intended to be used. In the second part of the argument, Warren considers cases where there is no intended usage to guide interpretation. Given the weakness of reference magnetism in the first set of cases, Warren claims that there is no good reason to think that reference magnetism is stronger in the second set of cases. Thus, reference magnetism is always arbitrarily weak.

To support the first part of his argument, Warren asks us to imagine that we have made contact with an alien civilization and we have discovered how to successfully translate their language into ours. However, when it comes to interpreting one particular arithmetical sentence, we seem to have made a mistake. We thought one of their symbols functioned the same as our symbol for addition, and so we translated one of their sentences as '987654321 + 123456789 = 1111111110' and marked it as true. But the aliens insist that the original sentence in their language is false. The correct "sum" of the number we translate as '987654321' and the number we translate as '123456789' should be the number that we translate as '5'. Details aside, the point is that we interpreted the alien's symbol as meaning *addition* but the aliens interpreted

their symbol as meaning something that resembles *quaddition*. Since this mathematical operation isn't quite the same as the *quaddition* operation developed by Kripke (see Kripke (1982): 8–9), let's give it a new name: *Warrenddition*.

Whose interpretation is right? Let's grant Warren's assumption that *addition* is a more metaphysically natural operation than *Warrenddition*. According to Warren, if reference magnetism exists then its existence should entail that the aliens mean *addition* rather than *Warrenddition*. But, according to Warren, the aliens do not mean *addition*; they mean *Warrenddition*. The aliens insist that they mean *Warrenddition* and their usage of the symbol has always been consistent with that insistence. Usage defeats reference magnetism.

Might reference magnetism still do something in cases that go beyond usage? Some numbers are enormous in the sense that all finite beings (human or otherwise) are unable to compute them.[2] Consequently, there is no history of intended usage to help determine whether '+' means *addition* or *Warrenddition* when applied to these enormous numbers. There are also no human or alien dispositions regarding what they would have intended to say had they applied '+' to these numbers. Perhaps reference magnetism works in this case to "fill in the gaps" and fix the meaning of '+' to the more natural *addition*.

Not so, says Warren. Warren compares the arithmetical case to a case of vagueness. The word 'bald' could be interpreted to refer to all and only those people who have zero hairs on the top of their head. In some sense, this interpretation would be the most metaphysically natural one. The word 'bald' does not mean that, though. The word 'bald' is semantically undecided, and this semantic indecision defeats reference magnetism. But the arithmetical case involving enormous numbers is also semantically undecided. According to Warren, if semantic indecision defeats reference magnetism in the case of vagueness then it should still defeat reference magnetism in the case of enormous numbers.

2 - Diversifying the Diet

In Warren's initial arithmetical case, the aliens insist that we've assigned their sentence the wrong truth value. But Warren claims that insistence is not essential to the case's significance. Warren also considers a variation of the case where the aliens are thoroughly uncertain: uncertain about whether their sentence is true or false, uncertain about whether our interpretation of them is the right one, and maybe even uncertain about whether their uncertainty is justified.

This thoroughgoing uncertainty is reminiscent of a cluster of cases given by Joseph LaPorte (2004). There, LaPorte wants to resist the standard story philosophers have inherited from Kripke and Putnam about the meanings of natural kind terms.

[2] For a more precise definition, see van Inwagen (1992): 143–146.

As the standard story goes, when we use the word ‘water’ we refer to the stuff that we drink on Earth but do not refer to the stuff that Twin Earthlings drink on Twin Earth because the word ‘water’ is supposed to refer to the stuff that we drink and whatever else has the same microstructure as that stuff. Since the stuff we drink has one chemical structure, H_2O, and the stuff they drink on Twin Earth has a different chemical structure, XYZ, the word ‘water’ refers to H_2O but does not refer to XYZ.

LaPorte claims that ‘water’ now refers to H_2O but, contrary to the standard story, it was not the case that ‘water’ referred to H_2O in the past. According to LaPorte, ‘water’ was semantically indeterminate in the past because ‘same microstructure’ back then didn’t mean what it means now. Nowadays, of course, we use the word ‘water’ to refer to all H_2O, including both *light water* (i.e. water containing the protium isotope) and *heavy* water (i.e. water containing the deuterium isotope). But that is because we resolved the semantic indeterminacy when it was discovered: we decided that ‘same microstructure’ meant ‘same atomic number’ rather than ‘same isotope’. Prior to our decision, though, there was no fact of the matter as to whether our use of ‘water’ referred to both *heavy water* and *light water* or referred to only *light water*. More generally, many words have open texture: their meaning is semantically indeterminate in a way that is hidden from use but becomes exposed by new information.

It seems that Warren has joined LaPorte’s side. Imagine a Deuterium Earth where *heavy water* plays the roles on Deuterium Earth that *light water* plays on Earth. Suppose that an Earthing tries to translate Deuterium English (Denglish) into Earthling English (English). The Earthling assumes that the Denglish word ‘water’ should refer to both *heavy water* and *light water* since the Denglish word ‘water’ refers to whatever has the same microstructure as that stuff that Deuterians drink. The Deuterians object strenuously and insist that the Denglish word ‘water’ refers to only *heavy water* because *light water* has a different microstructure. Warren should say that the Deuterians should be interpreted as they insist on being interpreted: in Denglish, ‘same microstructure’ means ‘same isotope’. Now suppose that the exact opposite happens: the Earthling says the Denglish word ‘water’ refers to only *heavy water* and the Deuterians insist that it refers to both. Warren should still say that the Deuterians should be interpreted as they insist on being interpreted: in Denglish, ‘same microstructure’ means ‘same atomic number’. In both cases, usage defeats the standard story about the meanings of natural kind terms.

I think all of this shows that there are important differences between the variations of Warren’s alien case, contrary to what he claims. In the original case, there is no semantic indecision. The aliens are aware of the different mathematical operations and they insist that *addition* is not the one that they intend to use. The variations where the aliens are uncertain, though, are underdescribed. If they were not aware of *addition* (like how students are not aware of *quaddition* before reading Kripkenstein), then our translation in effect exposes some open texture in their language. If that is where their uncertainty comes from, then the case should be treated like the Deuterium Earth case. If that is not where their uncertainty comes from, then I

have a hard time understanding why their uncertainty should be respected. They have all the relevant non-semantic information and they are not being given any new semantic information. Whence the uncertainty?

Warren may say that this final variation is similar to the vagueness cases. The word ‘bald’ has many plausible candidate meanings but ‘bald’ remains semantically indeterminate because we refrain from choosing one of those candidates. Similarly, the aliens are aware that they could mean *addition* or they could mean *Warrenddition* but they have refrained from choosing one or the other. Thus, Warren may say, if our indecision regarding ‘bald’ defeats reference magnetism then the aliens’ indecision should as well. I think this argument fails. But I won’t be able to explain why until later.

There is one last case that I want to present. Warren compares the metasemantic role allegedly played by reference magnetism to that of charity. So it seems reasonable to ask whether usage also defeats charity. Of course, what charity is, precisely, is a matter of some dispute. In what follows I will use Warren’s own Minimal Principle of Charity: “when interpreting, minimize the attribution of inexplicable errors. In general, *ceteris paribus,* the interpretation that minimizes the attribution of inexplicable errors is to be preferred” (Warren 2016: 186–187).

Suppose, as in Warren’s original alien case, the aliens object strenuously to our translation of the arithmetical sentence. Warren’s Minimal Principle of Charity would initially recommend that the aliens should be interpreted as meaning *Warrenddition*, since if we interpreted them as meaning *addition* then they would appear to be making an inexplicable computational error. But suppose that, after some further discussion, the aliens insist that they mean the same thing we do; they insist that they mean *addition*. We should conclude that they have made an inexplicable error. But we can push things further. Suppose that the sentence in question was the most computationally easy sentence possible and, thus, getting it wrong would be the most inexplicable of errors. Suppose that the aliens really do insist that they mean *addition* and suppose that they really do insist that the sentence is false when it should be true. We should conclude that the aliens have made the most inexplicable of errors. Usage also defeats charity.

3 - A Sketch of a Theory

Warren’s mistake is to assume that the metasemantic constraints operating in one context are the same metasemantic constraints operating in every other context, and thus the force of reference magnetism cannot change. But, as with all constraints (constitutive or otherwise), the force of reference magnetism can change.

In most ordinary conversational contexts, reference magnetism has no noticeable effect on what we say. Consequently, ordinary attempts to refer to the border of California cannot be taken hostage by a geological division no matter how much more natural that geological division is compared to whatever arbitrary division we intend to refer to. But in extra-ordinary

contexts the metasemantic constraints are different. Consequently, when Sider uses 'exist' he should be interpreted such that 'exist' means whatever available meaning is most natural, even if he is mistaken about what the most natural meaning of *existence* is and as a result what he says about it is wrong.[3]

This change in the force of reference magnetism is just one of the ways in which how we interpret what is said changes from context to context. Another change is exhibited by loose talk, where the utterance's communicated content differs from its literal content. When my friend says that Shanghai is 1200 kilometers from Beijing, it would be inappropriate for me to criticize them for saying something false (since Shanghai is slightly more than 1200 kilometers away from Beijing, depending on the route you take). Similarly, David Lewis speaks truly when he informs us that he looked in the fridge and there is no beer.

Warren may say that loose talk is a pragmatic phenomenon, insist on a firm distinction between what an utterance pragmatically communicates and what it semantically communicates, and claim that reference magnetism only exists if it affects what an utterance semantically communicates. He may then say that metasemantic constraints on what is semantically communicated cannot change from context to context. But the force of charity, which Warren accepts as a metasemantic constraint on what an utterance semantically communicates, also changes from context to context. When I appear to be disagreeing with someone, I am very likely to attribute the apparent disagreement to a miscommunication when that someone is a new student that I don't know well. In contrast, I am more comfortable attributing an error to my co-author because I know what they know, I know how they talk, and so on. This difference in interpretation initially appears to be in tension with Warren's Minimal Principle of Charity since my familiarity with my co-author makes their error even more inexplicable. The tension is resolved by recognizing that the force of charity changes: it really would be more uncharitable to interpret my co-author as making an inexplicable error, but in that context the force of charity is not particularly strong.

Changes in metasemantic constraints in some sense depend on us. We tend to interpret each other loosely when engaged in casual conversation but begin to tighten our interpretations when we think the details matter.[4] The courtroom, for instance, is an alethically demanding context that reduces the force of charity and quite regularly permits the attribution of inexplicable errors. The courtroom is an alethically demanding context because we make it so: that's one function of all the oaths and judicial procedures. Similarly, we make the ontology room a fidelically demanding context that increases the force of reference magnetism.[5]

---

[3] Sider gestures toward this response when he says that we should "stipulatively remove any normal metasemantic pressure toward tolerant interpretations that assign non-joint-carving meanings to quantifiers" (Sider 2011: 204).

[4] The literature on loose talk is large, but see Carter (2019) for a useful discussion.

[5] Perhaps it would be good practice for philosophers to recite an oath before entering the ontology room: "I swear to use joint-carving terminology and nothing but joint-carving terminology, so help me Lewis."

All of this should be contrasted with the metasemantic chaos endorsed by Herman Cappelen (see, *inter alia*, Cappelen (2018): 69–70). According to Cappelen, the metasemantic facts about what determines meaning are constantly in flux. Thus, even if reference magnetism is a powerful metasemantic constraint in the ontology room today, it might not be so tomorrow. As far as I can tell, Warren and I agree that there is some degree of metasemantic stability with respect to what the metasemantic constraints are and how they operate. But we seem to disagree about the extent to which these constraints are context-sensitive.

How this context-sensitivity operates is something that a complete metasemantic theory should explain in detail. In lieu of that, I will apply this sketch of a theory to the cases presented above.

Warren's original alien case is a perfect demonstration on how changes in metasemantic constraints depend on us. Generally speaking, insisting on the importance of one of the metasemantic constraints increases the force of that constraint and reduces the force of other constraints. So, when the aliens insist on their intended usage, the aliens increase the force of intended usage and reduce the force of reference magnetism. This general dynamic also explains my charity case: when the aliens insist on their intended usage, they reduce the force of charity and therefore should be interpreted in a way that entails they have made an inexplicable computational error. If they hadn't insisted, then we would have interpreted them more charitably.

One way to characterize the disagreement between LaPorte and those who accept the standard story about the meanings of natural kind terms is as a disagreement about the force of the metasemantic constraints in those contexts. Everyone should agree that *sameness of atomic number* is more natural than *sameness of isotope*. But LaPorte should also say that that difference in naturalness was not strong enough to resolve past interpretations, and thus the word 'water' was semantically indeterminate before the open texture was discovered. Those who accept the standard story should say that words like 'water' are typically used in contexts where the force of reference magnetism is made strong enough to automatically smooth out open texture, and thus 'water' was always based on *sameness of atomic number*. That being said, such a metasemantic policy doesn't have to guarantee complete semantic determinacy. In some cases, reference magnetism might fail to automatically smooth out open texture because the candidate meanings are equally natural. So, for instance, when Sider claims that he uses 'exist' to express the most natural sense of *existence*, his use of 'exist' would be semantically indeterminate if there were two perfectly natural modes of *existence* and neither of them was more natural than the other.

Warren argues that the case where the aliens are undecided about their arithmetical sentence is relevantly similar to cases where we are undecided about sentences involving vague words like 'bald'. According to Warren, the latter indecision leads to semantic indeterminacy, and so the former indecision should as well. Earlier, I said that this argument fails. It fails because the two contexts are importantly different. In standard cases of vagueness, the semantic indeterminacy

persists because we remain indecisive and we remain indecisive because deciding isn't important. Nothing of significance depends on where the precise border of 'bald' is drawn. In standard cases of open texture, though, a lot depends on how the open texture is smoothed out. *Heavy water* and *light water* behave quite differently, and so we need to decide which we intend to refer to when we use the word 'water'.[6]

It's hard to imagine why an alien civilization would remain indecisive about one of its basic arithmetical operations. If they are anything like us, they would want to resolve the indeterminacy between *addition* and *Warrenddition* because they would think having a determinate answer is important. This particular computation might never actually be required, of course. But the simple fact that it could be required should be enough to motivate a semantic decision in civilizations with interests similar enough to our own.

Some numbers are enormous, and consequently we cannot entertain semantic decisions regarding operations on these numbers. But that inability doesn't immediately entail that our words are semantically indeterminate. It may be that the standard story for natural kind terms should be extended to mathematical terms. If so, then in most contexts we increase the force of reference magnetism so as to automatically smooth out the open texture of our mathematical terms that, by definition, we cannot discover. I don't know if that's the right story to tell. But nothing Warren has said successfully speaks against it.

Wolfgang Schwarz, like Warren, argues against the existence of reference magnetism (Schwarz (2014): 29–31). Schwarz then proceeds to explain how naturalness plays a merely semantic role in determining the application conditions of some words. According to Schwarz, the word 'temperature' refers to the most natural thing that plays the thermodynamic temperature role because we use the word in that way and give it that meaning. For other words, naturalness is irrelevant (34). No part of this story, says Schwarz, requires a metasemantic constraint of reference magnetism.

I grant that naturalness sometimes plays the semantic role that Schwarz says it does. But Schwarz's alternative would struggle to accommodate the context-sensitivity under consideration. If I am right, then naturalness will be highly relevant to the meaning of one utterance and irrelevant to the meaning of another utterance even though the two utterances contain the exact same words. At first pass, then, fans of Schwarz's alternative would have to accommodate this context-sensitivity by saying that many words are semantically ambiguous between a "natural" reading and an "unnatural" reading. In contrast, fans of reference magnetism should say that language behaves more like the word 'flat' insofar as it has a single meaning with a context-sensitive standard. I side with the fans of reference magnetism here, although I admit that the final answer might depend on how the theory of reference magnetism coheres with broader theories of context-sensitivity.

[6] Of course, the decision may be made arbitrarily. But in this respect it is no different than a decision about which side of the road to drive on.

Conclusion

Warren's argument against the existence of reference magnetism seems to assume that the metasemantic constraints governing interpretation are not context-sensitive. But there are good reasons to think that they are context-sensitive. Given this context-sensitivity, Warren's argument succeeds in showing that reference magnetism is arbitrarily weak in some contexts but fails to show that it is arbitrarily weak in the contexts that matter.

References

Cappelen, H. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.

Carter, S. 2019. "The Dynamics of Loose Talk." *Nous* 55(1): 1–28.

Hirsch, E. 2009. "Ontology and Alternative Languages." In *Metametaphysics*, edited by David J. Chalmers, David Manley, and Ryan Wasserman, 231–259. Oxford: Oxford University Press.

Kripke, S. A. 1982. *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Laporte, J. 2004. *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

Schwarz, W. 2014. "Against Magnetism." *Australasian Journal of Philosophy* 91(1): 17–36.

Sider, T. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Van Inwagen, P. 1992. "There is No Such Thing as Addition." *Midwest Studies in Philosophy* 17 (1):138–159.

Warren, J. Forthcoming. "Reference Magnetism Does Not Exist." *Erkenntnis*: Online First.

Warren, J. 2016. "Internal and External Questions Revisited." *Journal of Philosophy* 113(4): 177–209.