Moral realism and semantic plasticity

David Manley
University of Michigan, Ann Arbor

Are moral terms semantically plastic—that is, would very slight changes in our patterns of use have shifted their meanings? This is a delicate question for moral realists. A 'yes' answer seems to conflict with the sorts of intuitions that support realism; but a 'no' answer seems to require a semantics that involves hefty metaphysical commitments. This tension can be illustrated by thinking about how standard accounts of vagueness can be applied to the case of moral terms, and also by considering how realists should respond to the Moral Twin Earth problem. After presenting the puzzle, I will argue that moral realists can accept the semantic plasticity of moral expressions while accounting for contrary intuitions in a way that is nearly cost-free.

1. Vagueness and plasticity

Consider some apparent cases of moral vagueness:[1]

i) A futuristic incubator contains a human sperm and egg. If no one intervenes, it will initiate conception and care for the developing organism until it is a human child of full moral status (whatever age that might be). At what point *exactly* would it be wrong for the owners of the original gametes to destroy the contents of the incubator? Likely this depends on further elements of the story not specified—but however we might do that it will seem odd to specify a precise threshold. After all, time can be sliced very thinly!

ii) If saving Jim requires destroying a single human cell sitting on a table, we should do it. And of course, had there been *two* cells on the table, that could hardly make the difference as to whether we should destroy what's on the table to save Jim. But enough differences of that magnitude—if they are the right ones—lead to a case where there is a human being on the table. At what point in this sequence of cases does it cease being obligatory to destroy what's on the table? (For those who consider the deep moral properties to be gradable, and the threshold at which

---

[1] For some discussions of moral vagueness interpreted as indeterminacy, see Railton 1992; Shafer-Landau 1994, 1995; Sosa 2001; Dougherty 2013; Wasserman 2013. On whether it is a different phenomenon from moral incommensurability, see Broome 1997, Wasserman 2003. iv) Wasserman considers cases where the continuity of personal identity is vague. (Imagine having made a promise to someone who then steps into a Parfittian teletransportation machine, so that it is vague whether that person comes out the other side.)

actions become obligatory to be shallow and context-dependent, we can still ask troubling questions about the underlying scale, such as: at what point does it cease being *better* to destroy what's on the table than to let Jim die?)

iii) I must choose between two outcomes: one improves certain lives along one axis of goodness—say, musical appreciation. The other improves those lives along another axis—say, bodily comfort. Now suppose we had a hyper-specific way of measuring these two goods, and outcome A involves an increase of precisely *m* units along the musical appreciation scale. It seems that for a range of precise points along the comfort scale, it is vague or indeterminate whether an outcome yielding that much comfort is better than A.[2]

Suppose we grant that such examples illustrate the vagueness of moral terms.[3] This can be unsettling for several reasons. If we hold that whenever is vague whether P, it is genuinely *indeterminate* whether P, we might find it unsettling that there could be no fact of the matter about whether one action is better or worse than another.[4] Or perhaps we hold that when it is vague whether P, there is some *unknowable* fact of the matter whether P. In that case we might find it unsettling that there are cases where one ought to perform an act, but there is no way to *know* that one ought to.[5]

For purposes of this paper, however, I'm interested in a different source of unease. The intuition I want to discuss arises when we apply theories of vagueness that involve *semantic plasticity*, the idea that slight differences in our use of certain terms can change their semantic values. This idea is a common commitment of theories that locate the source of vagueness in our *representations* of the world rather than in the world itself.

i) According to an important version of the *epistemicist* approach to vagueness, our use of a given vague expression like 'bald' succeeds in determining a completely

---

[2] Here the source of the apparent vagueness in 'better' is its 'multidimensionality': the fact that it involves giving weight to multiple potentially competing features of an outcome; see Schafer-Landau 1995: 84. However, I will set aside the question whether such cases are best thought of as incommensurability: see Chang 2002, Wasserman 2004, Broome 1997.

[3] We will consider one potential strategy for resistance below.

[4] Wasserman 2013 and Dougherty 2013 both appeal to this intuition in favor of the idea that moral realists ought to hold that moral properties are 'metaphysically important properties' (Wasserman 78) or 'part of the deep underlying metaphysical structure of the world' (Dougherty 7, 19). Though I focus on a different puzzle, the proposal I make in §7 may offer some comfort to more 'shallow' realists even on this score.

[5] For discussion see Dougherty 2013:10 and Constantinescu 2014: 23.

precise meaning, but very slight differences in how we use the term would have made a difference as to which meaning that is. As a result of this extreme sensitivity, we are not in a position to know which precise meaning is picked out. And it is our unavoidable lack of knowledge, rather than any semantic failure, that causes our befuddlement in Sorites cases.[6] Here the connection with semantic plasticity is obvious, since the view straightforwardly appeals to it.[7]

ii) According to *semantic indecision* accounts, vagueness arises from the failure of a linguistic community to settle on a precise meaning for an expression. There are various things a community can do to settle the meanings of expressions—use them in speech, form dispositions to use them, employ them inferentially in thought, and so on. But in the case of 'bald', we simply haven't done enough of those things. This leaves us with many closely related, equally good candidates for the meaning of 'bald', each of which we could have expressed had we been more precise in our usage. (Some—but not all—proponents of this sort of view also adopt a 'supervaluationist' approach to the truth-value of sentences containing 'bald'.)[8] Now, for us to exhibit the sort of care required to single out a precise semantic candidate for 'bald' may have required fairly significant changes from our actual pattern of use, which is marked by variability and ambivalence in borderline cases. But there are many ways in which we could have been this careful, differing only slightly *from one another*. For example, we could have consistently been disposed to use 'bald' to describe just those men with fewer than 382 hairs, or just those men with fewer than 383 hairs, etc. And if that difference was simply a matter of brute dispositions to use the expression—as opposed to an

---

[6] See chapters 7 and 8 of Williamson 1994.

[7] For discussion see Hawthorne 2006, Sennett 2012.

[8] For each sentence containing 'bald', and each precise candidate meaning for 'bald', we can consider whether that sentence would be true if 'bald' had had that precise meaning instead. On the supervaluationist approach, a sentence is true—or on another variation, determinately true—if it comes out true for all of these replacements, [determinately] false if it comes out false for all of them, and otherwise lacking [determinate] truth-value. For some broadly supervaluationist views see Fine 1975; Lewis 1982, McGee and McLaughlin 1995; Keefe 2000; Dorr 2003. For a discussion of the role of excluded middle in supervaluationism see Field 2000. However, the general approach of semantic indecision is open to other treatments of the truth-value of vague claims: for example, theories on which 'indeterminate' marks a truth value as opposed to a gap (see Field 2003), theories on which vague sentences are untrue (e.g. Braun and Sider 2007) or degrees-of-truth views (e.g. Weatherson 2005).

analytic association between 'bald' and any particular number of hairs—the difference between those two ways of using 'bald' might have been very slight.[9]

iii) Orthogonal to the question whether the vagueness of the public term 'bald' arises from community-wide semantic indecision or a kind of ignorance, we can also ask whether its content shifts due to facts that are local to a given occasion of utterance. The *contextual shift* approach helps explain why, at a context where we have accepted that a man with 100 hairs is bald, it is so difficult to deny that a man with 101 hairs is also bald. The idea is that assenting to the first claim activates a kind of principle of accommodation requiring us to interpret 'bald' in this context in such a way that anyone sufficiently similar to a man with 100 hairs also counts as bald. This helps explain a key phenomenon associated with vagueness: we have trouble identifying a precise boundary for the extension of 'bald' because, whenever we are in a context that involves focusing on a particular case, the boundaries move to accommodate that case along with a healthy buffer zone.[10] But the account certainly requires that vague terms be semantically plastic even from one context of use to another.

We come at last to the upshot. Take an act A that is a borderline case of wrongness. According to the version of epistemicism sketched above, A is

---

[9] To avoid this conclusion, one might insist that semantic precision of this sort would require a conscious and stipulative tie between (for example) 'bald' and 'having fewer than 383 hairs'. On this view, even consistent community-wide dispositions to use 'bald' to refer only to those with fewer than 383 hairs would not settle a precise meaning for 'bald' unless it was accompanied by a stipulative connection. (Note that the relevant claim must go beyond the mere idea that in most nearby scenarios where we succeed in meaning something precise with 'bald', we do so by way of a stipulation.) In that case, since arguably a shift in meaning-constitutive stipulations never constitutes a slight difference in usage, the semantic indecision view can avoid commitment to plasticity.

[10] For views that involve contextual shifts in the semantic content of vague expressions, see Kamp 1981, Raffman 1996, and Soames 1999: ch. 7. Delia Graff Fara's view is similar, but differs crucially with respect to semantic plasticity. On her view, a vague term expresses a single, interest-relative property in every context; but the extension of that property shifts from one context to another (2000; 2008). Thus 'bald' in a context expresses something akin to the property of being significantly balder than the relevant comparison class, as far as the speaker is concerned. Note that Fara implements her view within a framework for gradable adjectives where context settles a relevant degree or interval along an underlying scale that is *not* context-dependent. But arguably 'is bald' (for example) is not just vague along the axis of more-or-less hair; one must also consider patterns of distribution. Someone with plenty of hair on the sides and back of his head may be considered more bald than someone with fewer hairs evenly distributed across his head. Thus even the comparative 'is balder' is vague—at what point exactly does a tiny shift in distribution outweigh having one additional hair? Accounting for this vagueness requires deviating from the particular semantic implementation of her view that Fara gives.

determinately but unknowably wrong. It follows that 'A is wrong' is true in the actual world but false in a nearby world exactly like this one except that we use moral terms slightly differently there.[11] Meanwhile, the semantic indecision theorist will have to grant that there are pairs of worlds differing only with respect to minor use facts—worlds where we are more careful with 'wrong'— such that when we say 'A is wrong' we speak truly in one world and falsely in the other.[12] And the contextual shift approach, if it is to explain why we find it difficult to identify a precise boundary for a moral term, must appeal to the idea that its meaning shifts in order to charitably accommodate applications of the term in context.

The problem is that—especially for moral realists—it is counterintuitive to think that very slight facts in our use of moral expressions could shift which properties are at issue in our moral judgments. (By 'moral judgments' I mean judgments that play the same role in our lives that is actually played by judgments sensitive to moral facts.) Consider two of these scenarios in which our community has very slightly different dispositions to apply moral expressions and attitudes. Given semantic plasticity, it would be true for us to say in one of these scenarios, "If we had used moral expressions very slightly differently, we would not even be talking about wrongness with 'wrong', moral acceptability with 'moral acceptability', etc." Moreover, the semantic values of words are only part of a larger semantic package including the contents of our thoughts and of our other cognitive attitudes. So these slight usage difference would shift which properties govern the counterparts of moral intuitions in our lives. And members of the two communities would be expressing different motivations when they say things like "I want to do the right thing", motivations governed by different properties.[13]

---

[11] Wasserman 2013 points out this consequence of Williamsonian epistemicism.

[12] There are cases where it is tempting to think that the vagueness of a moral term is parasitic on semantic indecision concerning a descriptive term. For example, suppose I publicly promise a kiss on the head to every bald man who gives money to my charity. But one man who does so is a borderline case of baldness. What obligation, if any, do I have towards him (assuming he wants the kiss)? One might think that in this case the only semantic indecision involves 'bald', rather than 'ought' and related moral terms. But that is a confusion; nearby worlds in which I mean something precise by 'bald' are *not* at issue—the question at hand is whether I ought to kiss this man, given that I uttered the *vague* sentence 'I promise a kiss to every bald man who donates', with its actual meaning. The source of vagueness here is that the answer to *that* question is not settled by facts about how we use 'ought'. (Again, it may also be vague whether I promised to kiss men like him, but *given* that—what ought I to do?)

[13] These worries arise for the semantic indecision view under the counterfactual scenario where we use our moral expressions more carefully: we should then conclude that if we

None of this is to say that semantic plasticity commits the moral realist to anything like a 'social construction' or 'stance-dependent' view of the moral properties *themselves*. The fact that small changes in linguistic dispositions or conventions can induce a shift in which properties are expressed by a family of terms does not mean that the properties themselves are somehow constituted by the conventions or dispositions at issue. That would be to confuse semantic content with what settles the determination of that content.[14]

Instead, there is intuitive resistance to the idea that *which* of several properties would have been at issue could turn on subtle shifts in our dispositions to use moral expressions. It seems that a scenario in which we apply 'wrong' to a few different acts than we actually apply it to is a scenario in which we have slightly different (and perhaps false) *views* about what acts are wrong, not one where the *content* of our attitudes is shifted to accommodate our use. For example, if we could meet a linguistic community on another planet that used moral expressions in this slightly different way, we would take ourselves to have a slight disagreement with them, not to be talking past each other.

This last way of putting things will, for some readers, call to mind a very closely related point that arises in connection with the well-known Moral Twin Earth thought-experiment. In fact, as I will argue in the next section, our puzzle about moral vagueness arises from the very same intuition that generates trouble on Moral Twin Earth for certain strains of moral realism.

2. Moral twin earth

Putnam famously asked us to imagine a world where a kind of liquid other than H_2O plays all the roles that H_2O actually plays in our lives: it fills the lakes and streams, falls from the sky, and so on (Putnam 1975). Moreover, peoples' use of 'water' and associated dispositions are just like those of actual English speakers,

---

had refined our use in a slightly different way, our related attitudes would have had different properties as their objects.

[14] In considering the application of epistemicism to moral vagueness, Dougherty contrasts 'robust realism' on which moral properties are metaphysically joint-carving natural kinds, with 'stance-dependent' realism, on which 'ethical facts and properties obtain in virtue of our thoughts and practices.' But this dichotomy ignores forms of realism on which moral properties are neither stance-dependent nor particularly joint-cutting. Such properties are at issue in many cases of vague predicates, such as 'big' or 'soft' or 'heavy'. There is nothing 'stance-dependent' about the various precise candidate semantic values for 'big', such as the property of taking up more volume than 6.214 cubic feet. But neither are they particularly joint-carving.

say prior to the discovery of the chemical nature of water.[15] Considering this world, we realize that the Twins do not mean by 'water' what we mean; so Earthlings and Twins aren't disagreeing when they call different substances 'water'. (Thus, for example, 'Water is H_2O' expresses a truth in the mouth of an Earthling but would express a falsehood in the mouth of a Twin.) The point of the example was to show that 'meaning just ain't'— at least entirely—'in the head'.

Terence Horgan and Mark Timmons (H&T) have presented a partially analogous thought experiment for moral terms.[16] I say 'partially' analogous because it doesn't involve inconspicuously changing the environment and holding the use facts fixed, as Putnam's example does. There is no genuine water on Twin Earth, but there are genuine moral properties on Moral Twin Earth. Instead, the difference between Earth and Moral Twin Earth lies in the way that Moral Twins use moral terms like 'wrong', 'better', 'ought' and so on. For simplicity, we are to suppose that consensus has been reached on Earth as to which acts and outcomes satisfy those expressions— in particular, everyone is consequentialist in their deployment of moral terms. Meanwhile Moral Twins are uniform in their deontological use of the same terms—that is, they apply 'wrong' in a way that tracks the sorts of descriptive properties on which a deontologist thinks wrongness supervenes, and so on for the other moral terms.

H&T parlay this example into a challenge for naturalistic moral realism, the view that (i) 'there are moral facts, and these facts are objective rather than being somehow constituted by human beliefs, attitudes, or conventions', and (ii) moral predicates express properties of a sort that can be countenanced by a broadly 'naturalistic world-picture' (H&T 1992: 221-7, 244). On such a view, it would seem, which naturalistic properties are expressed by our moral terms is a function of our use of those terms along with the distribution of candidate naturalistic properties in the world. As an example of what this function might look like, H&T consider a view on which moral expressions have as their semantic value whichever properties best fill the role of causally regulating moral attitudes in

---

[15] This last hedge is for simplicity but is probably not required. Even holding fixed what I know about expert opinions, if I consider as actual a scenario where it turns out that all the watery stuff around here is XYZ instead of H_2O, I am inclined to conclude that there has been a vast conspiracy and it turns out water is not H_2O after all. This indicates that in such a scenario I would have been referring to XYZ. Likewise for Twins, if they live in a world where the experts say 'water is H_2O'. This is related to the 'robot cat' and 'blue lemon' cases from Putnam 1970.

[16] See Horgan and Timmons 1991; 1992a; 1992b.

humans (243-6).[17] This variety of realism would seem to predict that it's possible for this role to be filled by some other property, with a corresponding shift in the semantic value of moral terms.

Assuming that reference to moral properties is fixed in this way, H&T's thought experiment asks us to imagine that the relevant attitudes in Earthlings are causally regulated by consequentialist properties, while in Moral Twins they are causally regulated by deontological properties. So on the kind of moral realism we are considering, it should be that 'wrong', 'ought', and the like express different properties on Earth than they express on Moral Twin Earth, just as the referent of 'water' shifts from Earth to Twin Earth. And if members of the two groups were to meet, 'recognition of these differences ought to result in its seeming rather silly, to members of each group, to engage in inter-group debate about goodness,' i.e. about whether consequentialism or deontology is correct (1992: 166).[18]

But there's the rub. H&T claim the thought experiment fails to deliver this verdict: 'reflection on the scenario just does not generate hermeneutical pressure to construe Moral Twin Earthling uses of 'good' and 'right' as not translatable by our orthographically identical terms.' That is, 'any apparent disagreements that might arise between Earthlings and Twin Earthlings would be genuine disagreements– i.e., disagreements in moral belief and in normative moral theory, rather than disagreements in meaning' (1992: 165). On the kind of moral realism we are considering, the thought experiment leaves the distinct impression that we

---

[17] Here it would be important to specify what counts as a 'moral attitude' in a way that is independent of their semantics: e.g. use of 'right' and 'wrong' in speech and thought, disapprobation, motivations of a certain kind, etc.

[18] Must Moral Twins be located in a different possible world, or will a different planet suffice? Geoffrey Sayre-McCord has argued that they could even be on our planet: 'Right here on Earth it's not hard to find people and even whole communities whose use of moral terms reflect a sensitivity to natural features of a situation that others pretty much disregard' (McCord 288). Nick Zagwill concurs: 'the point seems to me to be just as strong for earthlings who differ in moral theory' (514). But this isn't so. The idea is supposed to be that what governs the meaning of moral terms is the functional property of being whatever property causally regulates moral attitudes in actual humans. People with idiosyncratic users of moral terms can not by themselves shift the meaning of the terms any more than a people who are disposed to call fool's gold 'gold' can shift the meaning of 'gold' all by themselves. Admittedly, there is a tacit assumption of widespread overlap among human societies about which naturalistic properties regulate the use of moral terms. But the example of 'earthlings who differ in moral theory' is not sufficient to undermine this. Whether Moral Twins are thought of as on another planet somewhere or in another possible world, their 'moral' attitudes are not supposed to count as among those whose causal regulation is relevant to determining which properties are expressed by our moral terms.

share a moral language with the Twins. Moreover, in the event that the two groups met, 'inter-group debate would surely strike both groups not as silly but as quite appropriate, because they would regard one another as differing in moral beliefs and moral theory, not in meaning' (1992: 166).[19]

3. Semantic stability

Can the moral realist simply reject plasticity—that is, deny that changes in use facts of the sort we have been considering would imbue moral terms with different semantic values? One way she might do so is by appealing to an idea that has become quite orthodox in contemporary metaphysics, namely that some properties do a better job of 'carving nature at the joints' or 'limning the structure of reality', while others are objectively more disjunctive, gerrymandered, 'gruesome'.[20] In David Lewis's terminology, the former are more *natural* than the latter; but since 'natural property' already has a different use in meta-ethics, I will use 'joint-carving property' for the Lewisian idea.[21] Crucially for our purposes, when

---

[19] One way to mitigate the effect of the thought experiment is to emphasize the significance of our *practical* disagreement with Twins (Merli 2002; Copp 2007: 214-16). We have clashing pro-attitudes that lead us to conflicting actions—meeting our Twins we might say 'That's not what we would do', and 'It's not right not to be motivated by that property'—and we'd be correct on both counts. (Meanwhile, the Twins will have a corresponding intuition about us, though their disapprobation will be *moral** rather than moral.) But is it plausible that, in the Moral Twin Earth example we are confusing a practical conflict (along with moral disapprobation) with moral disagreement? Horgan and Timmons, of course, would argue that we directly intuit moral disagreement in their example. But at any rate, the 'practical disagreement gambit' is not very helpful when we apply it to the case of vagueness. Focusing on our difference in use with our nearby counterparts, we may notice (for example) that they tend to apply 'wrong' slightly later than we do in the incubator case. Still this does not generate any sharp sense of disagreement: both we and our counterparts are presumably tentative when applying 'wrong' to cases that we sense are close to or at the borderline. So our sense that we mean the same thing that they do with our moral expression is not *derivative* on a sense of genuine disagreement. It just stems from the sheer strangeness of thinking that slight changes in our use 'wrong' could result in our meaning something different by it.

[20] See Dougherty 2013: 8-9.

[21] See for example Lewis 1983, Sider 2011, Dorr and Hawthorne forthcoming. There are various questions about this notion that we can set aside for our purposes, for example: is the relative notion at issue here primitive, or can it ultimately be understood in terms of what Lewis calls perfectly natural properties, along with more and more complex Boolean combinations thereof? And what exactly is the relationship between joint carving and

terms pick out fairly joint-carving properties, such as natural kind terms, they may not to be as semantically plastic as those that express metaphysically lightweight properties. There are two potential reasons for this.

i) First, we often intend to be picking out something like a natural kind as opposed to a collection of surface features, and this intention may trump other aspects of our use if there is no natural kind that perfectly fits the use facts. The relevant intention is likely to be implicit, but it may manifest itself in dispositions to apply the term in various scenarios. (For example, did speakers prior to the discovery of the molecular structure of water actually mean water by 'water', or did they mean watery stuff? One way to find out, if we were time-traveling experimental philosophers, would be to get their reaction to some thought experiments.)[22]

Of course, the intention to pick out a natural kind with a term need not be so authoritative that, in the absence of any candidate kind, the term will fail to refer. (More on that point later.) But if the metaphysical structure of the world cooperates, this kind of intention can impede semantic plasticity. Suppose there are many candidate semantic values that answer almost equally well to all other aspects of our use of a term *t*. Even so, if there is only one candidate natural kind, and we intend to refer to a natural kind, slight changes in other aspects of our use of *t* may not be sufficient to shift its meaning.

ii) The other way that joints can impede plasticity is through reference magnetism—at least if we follow Lewis in accepting such a phenomenon. The idea behind reference magnetism is that, even if we have no tacit intention to pick out joint-carving properties, they are just inherently more suitable as semantic values than others (Lewis 1983:375). In short, the world may play a more active role in the determination of semantic values than we might have supposed. In particular, where a semantics that emphasizes truth-oriented charity might assign extremely gruesome semantic values to certain terms, a better semantics might sacrifice some

---

various other notions in the neighborhood used by metaphysicians, such as grounding, fundamentality, and objective similarity?

[22] We could ask them to suppose it turns out that the watery stuff on Earth has a common underlying structure, and then test their intuitions as to whether a liquid on a distant planet with the same surface features but a different underlying structure would count as 'water'. Or we could ask them to suppose that as a matter of fact there is a common structure to 98% of the quantities they consider paradigms for 'water', but the rest is made up of a motley assortment of liquids with other kinds of structure. Suppose that they deny that the distant watery stuff in the first example is water, and intuit that the remaining 2% of watery stuff in the second example is only 'fool's water'. (See fn 14.) Then they likely intend 'water' to track an underlying explanatory kind of similarity whose nature is unknown, not a conjunction of surface features.

truth in order to avoid such semantic values. (This needn't be a magical, unexplained phenomenon: Williams 2007 argues that it is an interpretive constraint motivated by a general theoretical virtue of simplicity.)

Given all this, perhaps the moral realist can claim that moral properties have sufficient metaphysical heft to guarantee a high degree of semantic stability—whether because of reference magnetism, a tacit intention to pick out the most joint-carving kind available, or some combination of both.[23] But this view carries a significant theoretical cost, at least for the naturalistic realist. For one thing, given how different the use facts are between Earth and Twin Earth, this kind of account would seem to require that moral properties are *much* more joint-carving than any other semantic candidates, in order to guarantee that the same property is expressed on each planet. But from a naturalistic point of view, the properties on which the truth of moral talk actually supervenes—whatever those properties are—seem unlikely to be *any* more joint-carving than various nearby properties. Thus, for example, the naturalistic properties that a consequentialist might point to as constituting goodness seem no more or less joint-carving than the naturalistic properties a deontologist might point to. (This is especially clear if we assume a reductionistic metaphysics—both properties will then appear highly gerrymandered from the 'ground-floor' point of view—but it also appears true even if we assume, for example, that mental properties are irreducible.)

Few philosophers who consider themselves naturalists are likely to claim that moral properties are irreducible; but it is theoretically possible to divorce the idea of joint-carving properties from the idea of fundamentality. For example, one simply might give up on any hope of reductive naturalism and hold that there are 'emergent manifest properties, including moral properties, that command high [semantic] eligibility despite the gruesome nature of their supervenience base' (Hawthorne 2002:178). On such a view, moral properties may be highly joint-carving even though they appear highly disjunctive when considered from the

---

[23] Related views are endorsed in Hawthorne 2002, Wasserman 2013 and Dougherty 2013. Van Roojen (2006) appeals to a notion of 'discipline-relative' reference magnetism on which 'the kinds of properties which are more natural for the purposes of physics may not be the same as those which are more natural for purposes of biology'. The idea is that a property may count as joint-carving with respect to a discipline, and thereby acquire reference magnetism. The hope is that ethics itself could be such a discipline. But there is a problem of circularity here—van Roojen appeals to the idea that the best candidate semantic values for the terms of a discipline will be ones that best answer to its aims. But those aims are described in semantic terms, such as answering the 'questions posed' by the discipline, weighing 'evidence we have for different hypotheses', and so on (181). The problem is that which questions are being asked and which hypotheses are being considered will turn on which properties are assigned to be the semantic values of the discipline's expressions.

metaphysical ground floor. Unfortunately, granting an irreducible metaphysical halo to normative properties seems both theoretically costly and also in tension with most conceptions of naturalism.

Let's call a moral realist *hardcore* if she believes that moral properties are highly joint-carving, whether because they are non-naturalistic or irreducibly haloed. Otherwise, a realist is *mild*. As we have seen, the hardcore realist can simply vindicate the intuition of semantic stability illustrated by the Moral Twin Earth example. But what should she say about moral vagueness? Here she could deny that moral terms are vague at all—once she has non-natural properties in her ontology, perhaps it is not much additional cost to hold that they demarcate bright lines between acts that are extremely similar when considering only their naturalistic properties. Another option for the hardcore realist is to adopt a form of epistemicism that does not appeal to semantic plasticity. For example, perhaps in the case of 'bald' it is right to say that our irremediable ignorance about the boundaries arises from semantic plasticity—but perhaps in the moral case it arises from something different, viz. our inability to track the extension of the non-naturalistic property at issue.

Luckily, the realist need not accept any hardcore metaphysics. In the remaining section I will argue that mild realists can co-opt much of the benefit of the hardcore realist's account of semantic stability.[24]

4. Prospects for the mild realist

The (naturalistic) mild realist, it seems to me, must reject the deliverances of semantic stability intuitions in the Moral Twin Earth case. And if she wants to avail herself of one of the popular accounts of vagueness sketched in §2, she must also deny that moral terms are semantically stable in the vagueness cases, a result that seems to be in tension with moral realism.

Luckily there is more that the mild realist can say. Suppose she takes a page from the hardcore realist's playbook and claims that competence with moral expressions involves an (implicit) intention to express a joint-carving kind. Such an intention need not manifest itself as an iron-clad semantic constraint, where failure to satisfy the intention would induce outright reference failure. Witness 'jade': a paradigmatically natural kind term whose paradigms turned out to be made up of two very prevalent natural kinds with similar surface features. The community's reaction was to treat 'is jade' as expressing a disjunctive property—being either

[24] This is a different distinction than the one Dougherty draws between 'robust' and 'stance-dependent' realism. See my fn. 14 above.

nephrite or jadeite—rather than as failing to express any property at all.[25] In that case the intention to pick out a natural kind was one aspect of our use among many, that together formed the supervenience base for the semantic assignment. It was, we might say, a *provisional* referential intention.

To put things picturesquely, when the semantic gods assign semantic values, they must weigh a variety of factors. There is pressure to assign semantic values that yield a high proportion of true (or rational) beliefs, and satisfy as many referential intentions as possible. But presumably these must be weighted according to what we might call 'semantic authority'—for example, it may be that our intention to apply 'bachelor' only to males is sufficiently criterial that, holding our intentions fixed, 'bachelor' could not apply to a female regardless of how the world turned out. If so, our corresponding belief that all bachelors are male might be called 'analytic', though I would disown most epistemic associations with that term.[26] But the case may be quite different for, as it may be, the intention that 'jade' ought to pick out a natural kind. Likewise, take 'weight'—we intuit that it ought to pick out a unitary property, but we also intuit that people count as weighing less in low-gravity environments.[27] It turns out that no semantics can accommodate both of these intuitions; but we needn't hold that 'weight' is meaningless or give up on realism about weight. Perhaps it is vague whether 'weight' picks out mass or else some relation borne to the nearest planetary body. Or perhaps it picks out one and our provisional referential intentions in favor of the other are overridden.

The mild realist can grant that the semantic intention that moral terms shall pick out joint-carving properties is tacitly encoded in our competence with those terms. This is arguably why many philosophers have agreed that any properties perfectly answering to our conception of moral properties would have to be extremely strange. Hence Mackie:

---

[25] And if it had turned out that there were a hundred natural kinds evenly distributed among the stones called 'jade', perhaps our inclination would have been to treat 'is jade' as picking out a functional property—being a stone with such-and-such surface features—rather than a disjunctive-kind property with a hundred disjuncts. The jade example is discussed in connection with functional properties in Jaegwon Kim 1992.

[26] It is far from clear that one is ever in a position to know that a referential intention is iron-clad in this way. Errors might derive from a failure to be fully aware of our own dispositions, or even from a failure of imagination: many were surprised by Putnam's famous thought experiments to discover new epistemic possibilities, such as that 'Cats are animals' could in principle turn out false (Putnam 1970).

[27] See Field 2000 for discussion of the best candidate semantic value for 'weight'.

> If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. (Mackie 1977:38)

But—as we have seen with the cases of 'jade' and 'weight', we cannot assume that moral terms fail to pick out any properties unless every quasi-analytic belief about those properties arising from our referential intentions is borne out. In the absence of a perfect candidate, semantics may assign the best available alternative instead. In short, the intention to pick out joint-carving properties may be encoded in our competence with moral expressions; but it may also be *provisional*. If the world does not cooperate with a joint-carving semantic value, we needn't give in to an error-theoretic semantics. Given this view, it is hardly surprising that it is hard to shake the sense that moral predicates should express joint-carving, semantically magnetic propertie. Like a Müller-Lyer illusion, it is compelling even in the presence of a firm belief to the contrary.[28] This, I suggest, is the sort of phenomenon governing the intuitions of semantic stability in both the vagueness and the Moral Twin Earth thought experiments.[29]

In short, mild realism is perfectly compatible with the idea that we tacitly expect, due to our referential intentions, that moral properties are joint-carving. And this by itself can go a long way towards explaining the problematic intuitions of semantic stability at issue in this paper. Meanwhile, since this explanation needn't appeal to the kind of metaphysical baggage with which the hardcore realist is saddled, there is a case to be made that the mild realist has the better explanation.

---

[28] Matti Eklund (2005) and Jamie Tappenden (1993) have both broached the idea that there may be false claims that are analytic in the sense that finding them compelling is built into the competence conditions of an expression.

[29] It is hard to construct a fool-proof thought experiment to test for a tacit intention to pick out joint-carving properties, in part because that would require teasing apart intuitions of semantic magnetism. For example, we may consider as actual a scenario where the oracle tells us that overlapping very closely with the gerrymandered property of maximizing human happiness (or whatever naturalistic property best tracks our moral intuitions) there is also an emergent, highly joint-carving property. Even if its extension does not track our moral intuitions quite as well as the naturalistic property does, I intuit that in such a case we would take our moral terms to express the joint-carving property. But that may merely be an intuition of semantic magnetism.

Works Cited

Braun, David and Sider, T. 2007. Vague, so untrue. *Noûs* 41: 133–156

Broome, John. 1997. Is incommensurability vagueness? In Ruth Chang (ed.), *Incommensurability, incomparability, and practical reason*. Cambridge MA: Harvard University Press.

Constantinescu, Cristian. Forthcoming. Moral vagueness: a dilemma for non-naturalism. in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics* 9, 152-185

Copp, D. 2007. Morality in a Natural World. New York: Cambridge University Press.

Dorr, Cian. 2003. Vagueness without ignorance. *Philosophical Perspectives* 17(1):83-113.

_____ and Hawthorne, J. forthcoming. Naturalness. O*xford Studies in Metaphysics*.

Dougherty, Tom. 2013. Vague value. *Philosophy and Phenomenological Research*. 87:2.

Eklund, Matti, 2005. What vagueness consists in. *Philosophical Studies* 125:27-60.

Fara, Delia Graff, 2000. Shifting sands: an interest-relative theory of vagueness, *Philosophical Topics*, 28: 45–81. Originally published under the name 'Delia Graff'.

_____. 2008. Profiling interest relativity, *Analysis*, 68: 326–35.

Field, Hartry. 2000. Indeterminacy, degree of belief, and excluded middle. *Noûs* 34(1):1-30.

_____. 2003. No fact of the matter. *Australasian Journal of Philosophy*, 81: 457–480.

Fine, Kit. 1975. Vagueness, truth and logic. *Synthese* 54: 235–59.

Hawthorne, John. 2002. Practical realism? *Philosophy and Phenomenological Research* 64(1): 169-178.

_____. 2006. Epistemicism and semantic plasticity. In J. Hawthorne, ed., *Metaphysical Essays.* Oxford: Oxford University Press.

Horgan, Terence and Timmons, Mark. 1991. New wave moral realism meets Moral Twin Earth. *Journal of Philosophical Research* 16, 447-65.

_____ and Timmons, Mark. 1992a. Troubles for new wave moral semantics: the Open Question Argument revived. *Philosophical Papers* 21(3):153-75.

____ and Timmons, Mark. 1992b. Troubles on Moral Twin Earth: moral queerness revived. *Synthese* 92: 221-60.

Kamp, Hans. 1981. The paradox of the heap. In *Aspects of Philosophical Logic*, ed. U. Mönnich. Dordrecht: Reidel.

Keefe, Rosanna. 2000. T*heories of Vagueness*. Cambridge: Cambridge UP, 2000.

Kim, J. 1992. Multiple realizability and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52:1-26.

Lewis, David. 1982. Logic for equivocators. *Noûs*, 16: 431–441

_____. 1983. New work for a theory of universals. *Australasian Journal of Philosophy* 61: 343–77

2 J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin Books, 1977), p. 38. 16

McGee, Vann and Brian McLaughlin. 1995. 'Distinctions without a Difference'. S*outhern Journal of Philosophy* 33 (Suppl): 203–51.

Merli, David. 2002. Return to Moral Twin Earth. *Canadian Journal of Philosophy* 32(2):207–240.

Moore, G.E., 1903. Principia ethica. New York: Cambridge University Press.

Parfit, Derek. 1987. Divided minds and the nature of persons. In *Mindwaves*. Ed. Blakemore, C. and S. Greenfield. New York: Oxford.

Putnam, Hilary. 1970. Is semantics possible? *Metaphilosophy* 1(3):187–201.

_____. 1975. The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7:131-193.

Railton, P. 1992a. Pluralism, determinacy, and dilemma, *Ethics* 102: 720-42.

Sennett, Adam. 2012. Semantic plasticity and epistemicism. *Philosophical Studies* 161:273-285.

Shafer-Landau, Russ. 1994. Ethical disagreement, ethical objectivism and moral indeterminacy. *Philosophy and Phenomenological Research*, 54(2):331-344

_____. 1995. Vagueness, borderline cases and moral realism. *American Philosophical Quarterly*. 32(1):83-96.

Sider, Theodore. 2011. *Writing the book of the world*. Oxford: Oxford University Press.

Soames, S., 1999. *Understanding truth*, New York: Oxford University Press.

Sosa, E. 2001. 'Objectivity without Absolutes', in Byrne, Stalnaker, & Wedgwood (eds.), *Fact and Value: Essays on Ethics and Metaphysics for J.J. Thomson*, 215-227.

Tappenden, Jamie. 1993. The liar and sorites paradoxes: toward a unified treatment. *Journal of Philosophy*: 60 (11):551-577.

Wasserman, Ryan. 2004. Indeterminacy, ignorance and the possibility of parity. *Philosophical Perspectives* 18:391–403

_____. 2013. Personal identity, indeterminacy, and obligation. In Georg Gasser and Matthias Stefan (eds.), *Personal Identity: Complex or Simple?* Cambridge: Cambridge University Press.

Weatherson, Brian. 2005. True, truer, truest. *Philosophical Studies* 123 (1-2):47-70.

Williams, J. Robert G. 2007. Eligibility and inscrutability. *Philosophical Review* 116: 361–99.

Williamson, Timothy. 1994. *Vagueness*. London: Routledge.