

Normative reference as a normative question

Camil Golub

Forthcoming in *Erkenntnis*

Abstract

Normative naturalism holds that normative properties are identical with, or reducible to, natural properties. Various challenges to naturalism focus on whether it can make good on the idea that normative concepts can be used in systematically different ways and yet have the same reference in all contexts of use. In response to such challenges, some naturalists have proposed that questions about the reference of normative terms should be understood, at least in part, as normative questions that can be settled through normative inquiry. In this paper I have two goals. First, I argue that these naturalist proposals do not yet allow for radical disagreement on normative matters, or at least do not explain how such disagreement is possible. Secondly, I argue that, in order to account for radical disagreement, naturalists should not only treat normative reference as a normative issue but also adopt a non-representationalist account of normative concepts, on which such concepts are individuated through their practical role. I illustrate this point by showing how a view that combines naturalism and expressivism about normative discourse can vindicate the elasticity of normative concepts, their referential stability, and the objectivity of normative truths.

1 The metasemantic challenge to normative naturalism

Normative naturalism—the view that normative properties are identical with, or fully reducible to, natural properties—has many advantages. It is ontologically austere and fits into a broader naturalistic conception of the world, it trivially explains the supervenience of normative facts on non-normative facts, and it might even provide good answers to epistemological questions about normative thought, because it can rely on causal connections between normative beliefs and normative facts in explaining how it is that our beliefs are reliable.¹ Moreover, *non-analytical* versions of naturalism are arguably immune to G. E. Moore’s (1903) “open question argument,” and to similar objections to naturalism that trade on the different roles that normative and naturalistic concepts play in our cognitive and practical lives:² normative *concepts* have indeed a distinctive action-guiding function and cannot be defined in purely naturalistic terms, the non-analytical naturalist

¹See, e.g., Copp (2008), Locke (2014), Artiga (2015), Morton (2018).

²Two salient examples here are Parfit’s (2011) argument that, if naturalism were true, certain normative claims could not be substantive or informative, and Enoch’s (2011) argument that normative properties are “just too different” from natural properties for naturalism to be true.

can argue, and this helps explain the intuitions behind these objections to naturalism, but normative *properties* are just natural properties.³

However, naturalists have not yet properly addressed the challenge of providing an account of normative concepts and normative reference that allows for substantive and even radical disagreement on normative matters, while also making good on the objectivity of normative truths. Here are two examples that illustrate this challenge:

Moral Twin Earth Suppose that people on Earth systematically apply the term “wrong” to actions that fail to maximize utility. Now imagine a planet, Moral Twin Earth, where “wrong” plays the same role in guiding individual deliberation and interpersonal criticism, but the term is systematically applied to actions that violate Kant’s categorical imperative.⁴

Honor Code Imagine a community whose moral code is built around the preservation of personal honor and the elimination of impurity. This leads to moral judgments that radically diverge from our own judgments, but moral terms have the same normative role for this community as they do for us.⁵

Now, such examples have been used to argue against naturalism in various ways, so it might be more accurate to speak of a *cluster* of challenges here. For instance, Hare (1952) aimed to show that the descriptive meaning of moral terms is secondary to their evaluative meaning, and that people can disagree in virtue of shared evaluative meanings even if the descriptive meanings of their terms diverge. Horgan and Timmons (1991, 1993, 2000) argued that naturalism entails an objectionable kind of relativism in the face of Moral Twin Earth scenarios.⁶ More recently, Eklund (2017) has used a similar scenario to argue that naturalists and other realists cannot make good on the idea that reality itself favors our ways of valuing and acting, because it seems possible that different communities would use normative concepts with the same normative role but different extensions.

No matter which specific version of this challenge we focus on or how exactly we understand its goal, however, I propose that naturalists should aim to vindicate the following three theses in the face of examples like Moral Twin Earth and Honor Code:

Conceptual Elasticity Different people or communities can apply the same normative concepts to very different natural properties.⁷

Referential Stability Normative concepts that play the same normative role, *i.e.*, the same role in guiding individual actions and interpersonal criticism and advice, thereby have the same reference, even when applied to different properties.

Objectivity Normative properties are objective: they do not depend in the relevant sense on our beliefs, attitudes, or social practices.⁸

³See, e.g., Railton (2017) and Laskowski (2019).

⁴This is, of course, based on Horgan and Timmons (1991).

⁵I borrow this example from Williams (2018), but similar examples can be found in various challenges to naturalism, particularly in the work of expressivists like Hare (1952) or Gibbard (1990).

⁶In a similar vein, Streumer (2017: 49ff) argues that standard versions of naturalism entail a *false guarantee of truth* for judgments made under certain descriptively specified conditions.

⁷To clarify, this is not the claim that the same normative concepts can *refer* to very different natural properties for different people or communities. Indeed, this is what my next desideratum denies.

⁸I am relying here on a loose but intuitive notion of objectivity, and will not try to give a more precise

I will call the task of making good on these three desiderata *the metasemantic challenge to naturalism*.

To be sure, these theses are all controversial. In particular, some naturalists will reject the possibility of radical divergence in the use of the same normative concepts, as well as the related idea that Referential Stability holds true across a wide range of uses of such concepts. But I take these ideas for granted as desiderata for naturalism in the present paper. In other words, I will not try to make any new converts to Conceptual Elasticity, Referential Stability, and Objectivity. My goal is rather to explore whether and how naturalists can make good on these theses.

In response to examples like Moral Twin Earth and Honor Code, some defenders of naturalism (e.g., Sayre-McCord 1997, Copp 2000, Brink 2001) have adopted the strategy of treating questions about normative reference, at least in part, as normative questions that can be settled through normative enquiry. My first goal in this paper is to argue that these proposals take us in the right direction but do not fully make good on the three desiderata for naturalism, because they do not yet allow for radical disagreement, or at least do not explain how such disagreement is possible. My second goal is to argue that, in order to properly address the metasemantic challenge, naturalists should combine the idea of normative reference as a normative question with a non-representationalist account of normative concepts, on which such concepts are individuated through their normative roles. More precisely, I will argue that a view that I call *quasi-naturalism*, which combines naturalism with expressivism about normative discourse in a deflationary framework, can make good on Conceptual Elasticity, Referential Stability and Objectivity.⁹

2 The causal theory of reference for normative terms

First, I will discuss a metasemantic framework for naturalism which does *not* treat questions about normative reference as normative questions, namely Boyd’s (1988) causal theory of reference for moral terms, modelled after the causal theory of reference for natural kind terms proposed by Kripke (1972), Putnam (1975), and others. Diagnosing why this view is unable to make good on the three desiderata for naturalism will help motivate the different approach on which I focus in the rest of the paper, which puts normative theory at the center of metasemantic inquiry in metaethics.

For Boyd, the reference of a moral term like “wrong” is whatever natural property causally regulates its use in the right way. As Horgan and Timmons (1991) famously argued, this view seems to predict that “wrong” refers to different properties on Earth and Moral Twin Earth, despite playing the same normative role in both scenarios, so it

characterization of the relevant notion of dependence in this paper. Let me just say that objectivity thus defined is meant to rule out cases in which, say, what is morally wrong depends on moral attitudes or beliefs about what is wrong, while allowing that normative properties can depend in more benign ways on our mental states and social practices: for instance, that what is wrong can depend on facts about our hedonic states, and even on facts about relevant social institutions and conventions, as long as those institutions and conventions do not directly concern the given normative properties.

⁹This view is inspired by Gibbard’s (2003) argument that expressivists should be metaphysical naturalists, but unlike Gibbard I rely on a hybrid expressivist framework (more on this, in section 5). Bex-Priestley (forthcoming) defends a similar combination of hybrid expressivism and naturalism, which he also calls *quasi-naturalism*, but his primary focus is not on the metasemantic advantages of this view.

seems to fail to deliver Referential Stability. Moreover, according to Boyd, if two moral terms refer to different properties, then they encode different concepts. So his view does not seem to deliver Conceptual Elasticity either, and no sense in which the different communities genuinely disagree in cases like Moral Twin Earth.

Now, as Copp (2000) pointed out, this tight link between concepts and reference is not essential to the causal theory of reference as such. In principle, someone could adopt a causal account of reference determination and concede that terms like “wrong” refer to different properties in Moral Twin Earth cases, while claiming that the communities involved in those scenarios nevertheless use the same moral concepts.

However, this variation on Boyd’s view would still fail to deliver Referential Stability in Moral Twin Earth scenarios. As Horgan and Timmons (2000) put it, in clarifying their Moral Twin Earth challenge: naturalist views that rely on a causal theory of reference will entail either *chauvinistic conceptual relativism*, on which moral terms encode different concepts and have different referents in Moral Twin Earth cases, or *standard relativism*, on which moral terms encode the same concepts but refer to different properties in such scenarios. We are still nowhere close to a form of naturalism that vindicates Conceptual Elasticity, Referential Stability, and Objectivity for normative discourse.¹⁰

A more important objection to the argument that Boyd’s theory fails to meet the metasemantic challenge is that it ignores an important element of the view, namely what it is for a natural property to causally regulate the use of a term *in the right way*. The causal links with natural properties that determine the reference of moral terms, says Boyd, are those that tend to bring about true predication:

“Reference is ... an epistemic notion and the sorts of causal connections which are relevant to reference are just those which are involved in the reliable regulation of belief ... *Roughly* ... a term *t* refers to a kind (property, relation, etc.) *k* just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term *t* will be approximately true of *k*.” (Boyd 1988: 321)

This constraint on causal regulation is an instance of a more general epistemic access condition on reference, which Boyd and other naturalists accept. Here is how Boyd articulates this epistemic condition:

“It must be possible to show that our ordinary use of moral terms provides us with epistemic access to moral properties. Moral goodness must, to some extent, regulate the use of the word ‘good’ in moral reasoning ... The regulation need not be nearly perfect, but it must be possible to show that sufficient epistemic access is provided to form the basis for the growth of moral knowledge.” (1988: 328)

¹⁰Copp (2000) also suggests that, in cases where the same concept refers to different properties, there can still be *disagreement in attitude* involving that concept. Moreover, similar appeals to the idea in disagreement in attitude have been made by certain normative relativists and contextualists who try to account in this way for intuitive data about disagreement between people with different normative standards (e.g., Wong 2006, MacFarlane 2014, Finlay 2014). Arguing against these proposals goes beyond the scope of the present paper, but let me just say again that I am looking for an account that vindicates Referential Stability and Objectivity in cases like Moral Twin Earth and Honor Code.

“*t* cannot refer to *p* unless there are some people who, under ordinary circumstances, are at least pretty good at finding out about *p* and who reflect this capacity in what they say using *t*.” (2003: 515)¹¹

I will focus for now on Boyd’s application of this epistemic condition to the causal theory of reference, but the main worry I will raise applies to non-causal versions of the epistemic access condition as well.

Väyrynen (2018a) argues that the epistemic constraint on causal regulation might help account for sameness of reference in Moral Twin Earth cases, where the two communities disagree about fundamental moral principles but their particular moral beliefs are largely aligned and, we can assume, largely true. This seems to allow the naturalist to discriminate between the two candidates for the reference of “wrong”—the consequentialist property and the deontological one—by identifying which of them tends to bring about the true predication of wrongness in both communities.

However, Boyd’s view still fails to meet the metasemantic challenge, for two reasons.

First, this view, even if properly understood, still cannot account for sameness of reference in cases of radical disagreement like Honor Code, where it seems to predict either that the relevant terms refer to something else than our terms, or that they fail to refer to anything at all: for instance, assuming that Kantian deontology is the correct moral theory, one cannot plausibly claim that sufficient epistemic access to Kantian moral properties causally explains the true predication of wrongness in the Honor Code community, or indeed that there is much true predication of wrongness in that community.

The only way in which someone who adopts Boyd’s theory might try to account for sameness of reference in cases like Honor Code would be to claim that normative properties are multiply realizable and can be realized by very different natural properties in different scenarios. For instance, someone might hold that moral rightness is identical to the functional property of promoting social cohesion, and for us this functional property is realized by maximizing utility (or some similar property that causally regulates our use of normative terms), while for the Honor Code community it is realized by preserving honor and eliminating impurity.¹² But I believe this would be in effect to adopt a form of normative relativism, which sacrifices Objectivity.¹³

The other possible reply to the charge that Boyd’s view fails to allow for radical disagreement is to bite the bullet and insist that people in scenarios like Honor Code are not talking about the same normative properties as us.¹⁴ We quickly enter a stalemate here. While I have the strong intuition that there *is* genuine disagreement about what is

¹¹See also Boyd (1993) for further discussion on reference as an epistemic notion.

¹²Cfr. the evolutionary accounts of moral properties proposed by Sterelny & Fraser (2016) and Curry (2016). Boyd (1988, 2003) also suggests that moral properties might be realized by slightly different natural properties for different communities, depending on their history and specific social needs, but he does not endorse sameness of reference in cases of radical divergence in moral beliefs and practices.

¹³Again, I am relying here on an intuitive but imprecise notion of objectivity, which allows that the extension of normative properties can depend on relevant social practices (e.g., it can be an objective normative truth that I should follow local traffic laws, even if that means driving on different sides of the road in different countries) but does not allow for “right” to refer to the property of preserving honor and eliminating impurity in the Honor Code scenario.

¹⁴Merli (2002: 216) and Laskowski (2018: 722) suggest similar responses to Horgan and Timmons’ (1991) Moral Twin Earth argument and to Streumer’s (2017) related challenge to normative naturalism, respectively. See also Dowell’s (2016) argument that competent speakers’ semantic intuitions about disagreement have little probative force against externalist metasemantic theories.

wrong, good, etc. between us and the Honor Code community, and Referential Stability holds true in such cases, there is little I can say that will change anyone's mind on this issue. My goal in raising this worry about Boyd's theory is not to try to make progress in this dialectical standoff about radical disagreement, but rather to point out that a different approach, which treats questions about normative reference as normative questions, seems more promising as a way to make room for more disagreement on normative matters.

This is related to the second issue I want to raise about Boyd's theory: even in cases like Moral Twin Earth, where there is substantial overlap in moral beliefs between the relevant communities, it seems that the epistemic constraint on causal regulation will help account for sameness of reference only if we make substantive assumptions about what the normative truths are. However, Boyd insists that we should be able to settle facts about moral reference *without* relying on substantive moral commitments:

“It should be possible for non-human linguists, without anything like moral commitments, to investigate the semantics of human languages, including their moral components. Such investigators would plainly not be making moral judgments in any ordinary sense, so the naturalist faces the problem of indicating how they could determine the referent of, for example, moral uses of ‘good.’ ... I urge that a successful investigation into the metaphysics of morals could ‘in principle’ be carried out by the hypothetical extraterrestrials and that neither they—nor we, if amoralism is a psychological possibility for scientifically competent humans—would have to subscribe to, or accept, moral norms in order to carry out the investigation.” (2003: 533, 545)

In the same paper, Boyd sketches an account of how this kind of morally neutral investigation into moral reference is supposed to work: to fix the reference of natural kind terms, including moral terms, he says, we have to identify the reference candidates that best explain the inductive, explanatory, and practical achievements of the associated discourse, which we can do without making any substantive moral judgments.¹⁵

However, I do not see how this proposed methodology can work in the case of moral or normative terms, in particular how we can *correctly* identify the practical achievements of moral discourse in a given community without making substantive judgments about which social goals are in fact valuable or worth pursuing. If, on the other hand, we are not supposed to identify which goals accomplished by various communities constitute *genuine* moral achievements, but simply to defer to those communities' understanding of their own practical achievements, then again we will end up with a form of normative relativism, on which reference depends on parochial beliefs and social practices. No such relativist view can vindicate Referential Stability and Objectivity for normative discourse.¹⁶

¹⁵“In deciding what the referent is of a natural kind term, we are seeking to identify, from among the candidate categories, the one which best fits the explanatory role associated with natural kind terms: explaining the inductive, explanatory and practical achievements of the associated discourse. In a perfectly good sense of the term, we are making *normative* judgments here ... [But] the only normative judgments involved in the evaluation of semantic claims about natural kind terms are *epistemic* judgments about the cogency of competing explanations for achievements within particular domains of practice, and ... the only normative judgments which are implied by such semantic claims are hypothetical judgments about how to bring about such achievements.” (Boyd 2003: 535-536)

¹⁶Boyd does allow for the use of judgments about epistemic and hypothetical normativity in metase-

This is another motivation for a metasemantic approach that does not shy away from treating questions about normative reference, at least in part, as normative questions.

3 The normatively enriched causal theory of reference

The idea of normative reference as a normative question can be combined with a causal theory of reference for normative terms, and this is the strategy that Sayre-McCord (1997) adopts in the face of the problems faced by Boyd’s version of naturalism. The causal theory of reference can properly account for Moral Twin Earth cases, Sayre-McCord argues, if it is revised in one key respect:

“What a moral term refers to, if anything, is determined by whether, *in light of the best moral theory*, the use of that term can be seen as appropriately regulated by instances of a *normatively significant* kind.” (Sayre-McCord 1997: 291, my italics).

Thus, Sayre-McCord proposes a methodological shift in our thinking about moral reference: we should not defer to the natural or social sciences when isolating the general features of actions, states of affairs, etc. to which we refer when using moral terms. Rather, moral taxonomy and thus moral semantics should be driven primarily by moral theory.

This is not to say that science becomes irrelevant on this view: we need to examine whether the moral kinds identified by moral theory causally regulate our use of moral terms in the right way, which will be in part a scientific matter. But even here moral theory will play a key role: causal regulation of the right sort, Sayre-McCord suggests, involves a tendency to produce reliable beliefs that are responsive to new information about the given moral kinds. Thus, we need to rely on substantive moral assumptions in identifying which beliefs track moral truths and are responsive to relevant moral evidence.

This view seems able to account for Moral Twin Earth-type cases without major issues. Again, if two communities have different theoretical beliefs about morality but their particular moral judgments are largely aligned, the naturalist can argue, for instance, that the same moral kind causally regulates the use of “wrong” in both communities. This moral kind might be a Kantian property or a utilitarian property, or some other property that can only be identified by doing more moral theory. Moreover, in accounting for such cases, Sayre-McCord does not run into the self-imposed obstacles faced by Boyd, because he does not demand that we should be able to settle facts about moral reference by taking a purely external perspective to our use of moral terms, like the perspective of an alien linguist who has no moral commitments. Instead, we determine moral reference by doing both moral theory and empirical science.¹⁷

mantic inquiry (see the previous footnote). For instance: “if you want to achieve goal *X*, use *Y* to refer to *Z*”. However, this does not address the problem I raise here about how we can correctly identify the practical achievements of moral discourse in the first place. Boyd also concedes that, *in practice*, we will typically rely on our own moral judgments in assessing the aims of moral practice and therefore in settling facts about moral reference. But he insists that semantic inquiry into moral reference can in principle be “completely morally unengaged” (2003: 545), and this claim is the target of my worry here.

¹⁷See also Väyrynen (2018b: 207-208) for more discussion on the use of substantive normative assumptions in causal theories of reference for normative terms.

McPherson (2013) suggests that Sayre-McCord's proposal is still vulnerable to Moral Twin Earth-style challenges, precisely because it relies on substantive moral theory to settle facts about reference:

“Consider ‘substantive moral enquiry.’ On a non-normative gloss, this is perhaps roughly just what moral philosophers are now doing. It seems possible that such enquiry could be causally regulated by different properties at different places. On a normative gloss on ‘moral enquiry,’ however, the naturalist faces a renewed challenge to provide a non-Twin-Earthable naturalistic theory of reference for that concept.” (132)

However, I do not see the problem here for Sayre-McCord. First, on his view, moral kinds are determined by what is *in fact* the best moral theory, not by what some community or other takes to be the best moral theory. This is how Sayre-McCord can hold that the same moral kinds causally regulate the use of moral terms on Earth and Moral Twin Earth, even if the two communities have divergent theoretical beliefs and apply the expression “the best moral theory” to different natural properties. So the first horn of McPherson's dilemma does not apply to this view. As for the second horn, Sayre-McCord can simply apply his theory of reference to the expression “best moral theory” itself: more precisely, insofar as our use of the term “best” and the use of this term on Moral Twin Earth are both causally regulated in the right way by goodness as a moral kind, where the nature and extension of goodness are determined by what is in fact the best moral theory, then “best moral theory” will refer to that same theory for both communities, even if Moral Twin Earthers mistakenly take the term to apply to a different moral theory.

Perhaps the worry here is that we cannot rely on judgments about what the best moral theory prescribes in fixing the reference of “best moral theory,” because this will lead to an explanatory circle.¹⁸ However, I do not see anything objectionably circular about employing our own concepts in substantive ways when fixing the reference of those very concepts. Perhaps this kind of metasemantic explanation will only confirm what we already believed about the reference of those concepts, but this does not mean that there is no genuine explanation being provided here, or that a theory like Sayre-McCord's does not work for the very normative concepts on which it relies.

A different way to raise this circularity worry would be to focus on the broader theoretical goals of naturalism in metaethics: the promise of naturalism is to explain the nature of normative facts and properties, and the workings of normative thought and language, in fully naturalistic terms. Using irreducible normative notions in an account of normative reference would undermine this core naturalist goal, someone might argue.¹⁹

However, naturalists need not eschew the use of normative concepts in determining the reference of normative terms, or at least *non-analytic* naturalists have nothing to worry about here: if we accept that fundamental normative notions cannot be defined in purely naturalistic or descriptive terms, and that questions about normative reference are inextricably linked with normative theory, then we should not be surprised that irreducible

¹⁸Streumer (2017: 57) raises a similar regress worry about naturalistic views that rely on normative notions in specifying reference conditions for normative terms.

¹⁹Thanks to Nick Laskowski for suggesting this way of pressing the circularity worry.

normative notions show up in a naturalist account of normative reference, nor should we take this to go against the broader goals of the naturalist project. What matters for the success of this project, from a non-analytic naturalist perspective, is to reduce normative facts and properties to natural facts and properties, and to explain why natural creatures like us would have a need for irreducible normative concepts. Sayre McCord's normatively enriched causal theory of reference does nothing to undermine either of these goals.

My objection to this theory of reference concerns again radical disagreement: this view cannot account for sameness of concepts and reference in cases like Honor Code, where it is implausible that the same normative kinds causally regulate the use of normative terms in both communities.

Here too, one option for the naturalist would be to adopt a loose conception of moral kinds, on which, for instance, moral wrongness is a functional property that can be realized by very different properties for different communities depending on their moral beliefs, attitudes, or social practices. But this would mean abandoning Objectivity for moral discourse. Moreover, this option seems even less promising for someone who adopts a normatively enriched causal theory of reference than it did in Boyd's case: it is implausible that our best moral theory will establish that moral kinds like goodness, wrongness, etc. are radically multiply realizable and dependent on social practices and beliefs.

To be clear, this is not the route taken by Sayre-McCord in the face of this worry about radical disagreement. Instead, he rejects the intuitive data about disagreement and sameness of reference on this objection relies. Here is what he says about alleged shared reference in cases of radical disagreement:

“If we discovered of a community that their use of the terms ‘right’ and ‘good’ were not appropriately regulated by what is right or good but instead by something else we should again have grounds for thinking that they were not using the terms to say of things what we say with ours—even if their terms played a role in guiding their actions. And again we would presumably have a disagreement with them, yet it would be a mistake to see our terms ‘right’ and ‘good’ and their orthographically identical terms ‘right’ and ‘good’ as equally deployable in a discussion of what is right or good.” (1997: 289-290)

Moreover, Sayre-McCord suggests that our very intuitions about shared reference hinge on thinking that the uses of the relevant moral terms are causally regulated by the right things in the world (1997: 287).

I can report that *my* intuitions about shared reference in cases like Honor Code are not weakened by the belief that concept users in those scenarios are not responding to the right things in the world in their moral judgments. But again, I am not hoping this appeal to intuition will change anyone's mind here. My goal is to set out why even a normatively enriched version of the causal theory of reference cannot hope to make good on the three desiderata for naturalism that I take for granted in this paper.

Now, in the same passage cited above, Sayre-McCord suggests that we can still disagree with communities with radically different moral beliefs, despite the fact that our moral terms are not mutually translatable and do not refer to the same properties, and later he clarifies that this will be a kind of conflict in attitude: we think that those communities

should be concerned with what is good, wrong, etc., while they do not care about these moral properties (1997: 290, fn. 27). However, if this is meant to amount to genuine propositional disagreement about what moral kinds we should care about, Sayre-McCord’s view will still struggle to account for such disagreement, because it is unlikely that the same moral or normative kind causally regulates the use of “should” in both communities. If instead Sayre-McCord has in mind a *mere* disagreement in attitude that does not also amount to propositional disagreement about what we should care about, then again I believe this suggestion does not do justice to the intuitive data about disagreement here.

To sum up, we cannot fully account for Conceptual Elasticity, Referential Stability, and Objectivity as long we preserve a causal constraint on normative reference and concept use: there is no plausible conception of normative kinds on which the same properties causally regulate the use of normative concepts in all the relevant scenarios.

Moreover, the problem here goes beyond causal theories of reference for normative terms, and affects any view that adopts an epistemic access condition on reference.²⁰ What scenarios like Honor Code show is that whether someone uses a normative concept that refers to a certain normative property should not depend in any way on how reliable they are in their judgments about that property, or in their normative judgments more generally.

4 Referential intentions

A different metasemantic framework for naturalism that also treats questions about normative reference as normative questions seems to avoid the problems that I have raised so far. This is the kind of view proposed by Copp (2000) and Brink (2001), on which the reference of moral predicates is determined by certain referential intentions that can be ascribed to all users of such predicates, irrespective of their moral beliefs, and by substantive moral theory. I will focus on Brink’s proposal in this paper, but my arguments can be extended to Copp’s view as well with only minor changes.

According to Brink, anyone who uses a moral term like “right” intends to occupy *the moral point of view* in doing so, which means using such terms to pick out those properties that make actions interpersonally justifiable. This is how we identify who is using the

²⁰This includes, for instance, van Roojen’s (2006) proposal, which imposes a knowledge condition on normative reference. For similar reasons, I believe cases of radical disagreement cannot be properly accommodated by Williams’ (2018, 2020) substantive radical interpretation view, on which the correct interpretation of the content of an agent’s thoughts or claims is the one that does the best job of making her substantively rational, where substantive rationality is a measure of how well the agent is responding to normative reasons in her beliefs, attitudes, and actions. At first glance, this view also seems to predict that people in Honor Code-type scenarios do not refer to the property of wrongness in their use of “wrong,” insofar as their moral judgments are not formed in response to actual instances of wrongness, and therefore are not reason-responsive in this sense. Williams rejects this apparent implication of his view, and tries to accommodate such cases by focusing on other ways in which the relevant agents are being rational, e.g., in the links they exhibit between moral judgments and certain behavioral dispositions, like being disposed to blame people for actions one believes to be wrong. But on his view the fact that agents in Honor Code-type scenarios are vastly mistaken in their moral judgments and attitudes still seems to count as *evidence* against attributing the same reference to their concepts as to ours, even if this evidence might be outweighed by other considerations. I disagree: the fact that the relevant agents exhibit certain motivational tendencies in their use of moral concepts is the only thing that matters in establishing that they use the same concepts as us—and with the same reference, if this is what first-order moral theory tells us about these cases—no matter how unreliable those agents might otherwise be in their normative judgments and attitudes.

same moral concepts as us: by identifying who has the relevant referential intentions. And the reference of moral terms will consist in those natural properties that do in fact make actions interpersonally justifiable, where this is to be settled through moral theory.

This view seems to allow in principle for radical moral disagreement because it does not require that people be even remotely reliable in their moral beliefs in order to use moral concepts that refer to certain moral properties. For example, Brink can claim that the Honor Code community uses the same concept of rightness as us, in virtue of intending to occupy the moral point of view in using the term “right,” even though they are radically mistaken about which properties make actions interpersonally justifiable. Moreover, moral theory can vindicate Referential Stability and Objectivity with respect to such cases, if it establishes that the same objective natural properties make actions interpersonally justifiable in all the relevant scenarios.

Before raising my worry about this view, let me address two other objections to it.

First, Horgan & Timmons (2000) have argued that the normative notions that this type of view builds into the content of referential intentions, e.g., *interpersonally justifiable*, are too broad to deliver a determinate reference for moral predicates: for instance, both consequentialist and deontological properties of actions can fit the bill here.²¹

However, on this metasemantic picture, reference is determined by what is in fact the best moral theory, not by what some community or other takes to be the best moral theory, and it is reasonable to expect that the best moral theory will deliver fairly precise answers to the relevant moral questions. In particular, the fact that different communities might disagree about what makes actions interpersonally justifiable does not entail that moral reference will vary between those communities according to Brink. If, say, act utilitarianism turns out to be the best moral theory, then the objective reference of “right” will be the property of maximizing utility, not just for us but also on Moral Twin Earth, in the Honor Code scenario, and for any community that uses the same concept of rightness, no matter what their theoretical beliefs about morality might be.

To be sure, scenarios in which different communities systematically disagree in their moral beliefs might still pose an important challenge to Brink’s view. But this will be an *epistemological* challenge from disagreement, applicable to all versions of normative realism, not a metasemantic challenge about how we fix moral reference if we rely on the verdicts of the best moral theory. If we set aside such epistemological issues, and grant that Brink is entitled to rely on substantive theoretical moral beliefs even in the face of actual or possible disagreement, there is no lingering challenge here about whether his view can deliver determinate answers to questions about moral reference.

Secondly, someone might object that not all users of moral predicates will have the specific referential intentions that Brink attributes to them. In particular, it seems possible that a community would use moral concepts in ways that cannot be justified to all its members, and deliberately so: for instance, a community that openly excludes the interests of certain social groups in its moral norms and practices.

However, I believe Brink can deal with this worry as well. One way to do this would be

²¹Horgan and Timmons focus on Copp’s (2000) proposal and its reliance of the notion of human flourishing, but I take it that their objection is more general and would extend to Brink’s view as well.

to qualify the description of the relevant intentions: e.g., perhaps the referential intention associated with “right” is to pick out those properties that make actions *justifiable to all persons with full moral status*, or something along these lines. Another option would be to define the content of these intentions in even broader terms. For instance, the Brink-style naturalist might simply claim that all users of “right” intend to pick out *those natural properties that make actions right*. This might not sound like a very illuminating characterization of these referential intentions, but it would still allow the view to accommodate radical disagreement and to deliver a determinate reference for moral terms with the help of substantive moral theory. (Moreover, this way of understanding referential intentions would have the virtue of being applicable to non-moral normative concepts like *reason* or the all-things-considered *ought*, where it is less likely that referential intentions will involve the notion of interpersonal justification.)

My worry about Brink’s view is that, even if it *allows* for Conceptual Elasticity and Referential Stability to be true, including in cases of radical disagreement, it still does not fully *explain* how radical disagreement is possible. Here is why.

First, note that Brink takes the relevant referential intentions to have irreducible normative content:

“This account of moral semantics . . . is fiercely nonreductionist. To characterize the moral point of view in terms of interpersonal justification is to characterize it in ineliminably normative terms.” (2001: 176)

Moreover, any view with this structure that wants to allow for radical disagreement *must* claim that the relevant intentions have irreducible normative content. If the content of these intentions was specified in fully non-normative terms, the naturalist would again struggle to accommodate cases like Honor Code: it seems impossible to find a descriptive content for referential intentions that could be attributed to all users of normative concepts, no matter how much they diverge in their normative beliefs and practices.²²

However, if these referential intentions have irreducible normative content, we still have not been offered an account of what it is in virtue of which people on Moral Twin Earth, in the Honor Code community, and other relevant scenarios have referential intentions *with the same normative content* as our corresponding intentions, and therefore we still lack an explanation of why these people use the same normative concepts as us. For instance, what makes it the case that someone who intends to use “right” to pick out what they would describe as *features that make actions interpersonally justifiable* is employing the same concept of *interpersonally justifiable* that we do, and thus the same concept of rightness? Until such questions are answered, Brink’s view does not fully address the metasemantic challenge, because it does not explain the elasticity of normative concepts.

²²Take the claim that the referential intention associated with the use of “right” is to pick out those actions that are D, where ‘D’ is a stand-in for whatever descriptive feature might seem plausible in this context, such as maximizing utility or promoting social cohesion. We can easily come up with a counterexample to this claim: some actual or possible community that uses “right” (and the associated concept of rightness) without the intention of picking out actions that are D. If ‘D’ stands for maximizing utility, for instance, both Moral Twin Earth and Honor Code will do for this purpose.

5 The quasi-naturalist solution

In addressing the metasemantic challenge, both the strategy of treating questions about normative reference as normative questions and abandoning the epistemic condition on reference are steps in the right direction, but they are not enough. Naturalists still need to provide an account of what makes it the case that people who vastly diverge in their normative beliefs and practices can nevertheless use the same normative concepts.

In order to achieve this goal, I propose that naturalists should adopt a view on which normative concepts are individuated through their normative roles, rather than through representational relations with the normative realm. Naturalists can thus explain Conceptual Elasticity, while also vindicating Referential Stability and Objectivity by treating all questions about normative reference and objectivity exclusively as normative questions. More precisely, I will argue that expressivism about normative discourse can help naturalists fully address the metasemantic challenge. I will use the term *quasi-naturalism* for this combination of expressivism and naturalism about normativity.²³

The idea that expressivism is compatible with normative naturalism is not new. Gibbard (2003) famously argued that expressivists should accept naturalism about normative properties because any coherent planner is committed to there being a natural property that constitutes what is good, what one ought to do, etc. More recently, Bex-Priestley (forthcoming) has made a similar argument for naturalism in a hybrid expressivist framework and also uses the label *quasi-naturalism* for the resulting view.²⁴ But the advantages of quasi-naturalism in dealing with metasemantic challenges have not been fully explored yet. The main contribution of the present paper is to fill this gap.

Expressivism is, broadly speaking, the view that normative claims express desire-like mental states, such as plans or attitudes of approval and disapproval, where this is understood as a claim about normative meaning, and not merely about the pragmatics of normative discourse. While any expressivist view would be well-placed to explain the possibility of radical normative disagreement and make good on the three desiderata for naturalism, I will rely here on a hybrid version of expressivism, according to which normative claims express both desire-like attitudes and representational beliefs (Ridge 2006, 2007, 2014; Toppinen 2013). I find hybrid expressivism to be independently plausible, among other reasons because it can address the cluster of challenges to expressivism known as the Frege-Geach problem. Moreover, as Bex-Priestley (forthcoming) has argued, this version of expressivism can provide a clear interpretation of the metaphysical claims of normative naturalism, so it is particularly useful in addressing the metasemantic challenge to naturalism.

Let me introduce a simple version of hybrid expressivism, which will do for the purposes of explaining how expressivism can accommodate metaphysical naturalism and help vindicate Conceptual Elasticity, Referential Stability, and Objectivity. On this view, an

²³Expressivism is the most salient non-representationalist account of normative discourse, but it need not be the only option for naturalists: for instance, inferentialism about normative concepts might also do the job. But my goal here is only to defend the quasi-naturalist response to the metasemantic challenge.

²⁴Ridge (2014: 42, 132) also suggests that his version of hybrid expressivism is compatible with naturalism about normative properties. See also Copp (2001, 2018), Schroeder (2014), and Laskowski (2019, 2020) for other attempts to combine naturalism with expressivism or non-cognitivism, but with significantly different goals and commitments than quasi-naturalism as defined in this paper.

atomic normative sentence like “Genocide is wrong” expresses a complex mental state consisting in (1) an attitude of disapproval of actions that have a certain property, and (2) a belief that genocide has that property.

It is important to note that these complex mental states expressed by normative claims are multiply realizable (this is why some proponents of hybrid expressivism call them *relational* or *higher-order* mental states): they can be realized by different combinations of desire-like attitudes and corresponding beliefs for different speakers. For example, “Genocide is wrong,” when uttered by a utilitarian, might express an attitude of disapproving actions that do not maximize utility and the belief that genocide fails to maximize utility, while an ethical egoist’s utterance of the same sentence might express disapproval of actions that are not in the speaker’s best interest and the belief that genocide is not in the speaker’s best interest. But both utterances will express the same relational mental state: an attitude of disapproving of actions that have a certain property, combined with the belief that genocide has that property (Schroeder 2013, Toppinen 2013, Ridge 2014).

Here is how this hybrid expressivist view is compatible with naturalism, and with genuine objectivity for normative truths.²⁵

First, expressivists can endorse many tenets of normative realism, by relying on deflationism about truth, fact, and other related notions. For instance, on a deflationary account of truth, “It is true that p ” is equivalent to p , and this schema fully captures the meaning of “true”. Therefore, expressivists can hold that “It is true that genocide is wrong,” taking this claim to simply rehearse the normative judgment that genocide is wrong. Similarly, expressivists can claim that there are normative facts, or that normative judgments describe such facts, by relying on deflationary accounts on these notions.

Secondly, expressivists can endorse claims about objective normative truths and facts, by taking such claims to express a particular kind of attitude. Our attitudes toward certain actions or states of affairs are *resilient*: they remain the same even when we consider scenarios in which our attitudes were different. For instance, we disapprove of child abuse even when we consider a possible world in which we ourselves did not disapprove of it. According to an expressivist account of objectivity claims, it is such a resilient attitude that we express when we say that child abuse is objectively wrong, no matter what anyone thinks or feels about it.²⁶

Both of these points are familiar quasi-realist fare (e.g., Blackburn 1993, Gibbard 2003), and do not depend on a hybrid expressivist framework. But when it comes to accommodating metaphysical naturalism, hybrid expressivism becomes important.

Take the following naturalist thesis: “Moral wrongness is identical with failing to maximize utility.” On the simple version of hybrid expressivism introduced above, this claim

²⁵There is a long-running debate about whether and how we can distinguish between quasi-realist expressivism and genuine realism about normativity. In previous work, I have argued that there is no meaningful divide between quasi-realism and anything recognizable as a general notion of realism in metaethics (see Golub 2017, 2021), but I am setting aside this issue in the present paper. Here I am only claiming that expressivism can make good on the objectivity of normative truths, facts, and properties, whether or not this amounts to vindicating a form of normative realism.

²⁶More precisely, in a hybrid-expressivist framework, “Child abuse is objectively wrong” will express an attitude of disapproving of actions that have a certain natural property, even when considering scenarios in which we or others did not disapprove of such actions, and the belief that child abuse has that natural property.

can be understood as expressing an attitude of disapproving actions that have a certain property, and the belief that the property in question is identical with failing to maximize utility. And similar hybrid expressivist accounts can be given for naturalist theses about property reduction, constitution, or other metaphysical relations that fall short of identity (see again Bex-Priestley forthcoming). More generally, hybrid expressivists can offload the content of such metaphysical naturalist claims onto the content of the representational beliefs that are expressed by normative claims on their view.

In contrast, a pure expressivist view such as Gibbard's (2003) will struggle to account for the metaphysical claims of naturalism. As Bex-Priestley argues, such views seem to lack the resources to distinguish between naturalist identity or reduction claims, on one hand, and mere correlation claims involving normative and natural properties, on the other: for instance, "Goodness is reducible to being pleasurable" and "Necessarily, things are good if and only if they are pleasurable" would seem to express the same mental state according to Gibbard's plan-expressivism and other versions of pure expressivism.

To sum up, quasi-naturalism is the combination of hybrid expressivism and metaphysical naturalism in a deflationary framework, and the hybrid element is arguably necessary for this combination to fully work.

This view can easily explain the elasticity of normative concepts. For instance, quasi-naturalists will argue that judgments about what is morally wrong have a distinctive role in guiding individual deliberation and interpersonal criticism and advice precisely because they encode desire-like mental states, and both we and the communities in Moral Twin Earth and Honor Code use the same concept of wrongness in virtue of this shared attitudinal content that is constitutively linked with the normative role of "wrong". Thus, quasi-naturalists can allow for wide divergence in the use of normative concepts as long as the normative role of said concepts remains the same.

Quasi-naturalism can also vindicate Referential Stability and Objectivity for normative discourse, by treating questions about the reference of terms like *right* and *wrong* exclusively as normative questions, to which normative enquiry can establish unique and objective answers. Here deflationism is again a key tool. For instance, the question "What is the reference of 'wrong,' for us and on Moral Twin Earth?" will be treated as equivalent to "What kind of actions are wrong, for us and on Moral Twin Earth?". To answer this question, quasi-naturalists will argue, we need to identify the natural features that make actions wrong in these various scenarios, which we can only do through substantive moral enquiry. And again, expressivists have the resources to claim that the same features make actions wrong in all the relevant cases, no matter how any community uses the concept of wrongness: they can make sense of the idea of objective wrongness, objective goodness, etc., by identifying a special kind of attitude expressed by objectivity claims, like an attitude of disapproval of actions that have certain natural features even with respect to scenario in which we or others had different attitudes. By adopting such resilient attitudes, quasi-naturalists can coherently claim that terms like *wrong* and *good* have a unique objective reference.²⁷

²⁷To be clear, the quasi-naturalist account of Conceptual Elasticity, Referential Stability, and Objectivity that I have sketched here does not depend on a hybrid expressivist framework. Pure expressivists such as Gibbard (2003) will also tie the identity of normative concepts to their normative roles and will

6 Objections and replies

The quasi-naturalist response to the metasemantic challenge relies on tying the identity conditions of normative concepts to their normative roles: normative concepts are identical if and only if they encode the same desire-like mental states, and it is these desire-like mental states that explain the normative role of such concepts, that is, the patterns of causal connections between uses of normative concepts and various motivational dispositions manifested in individual deliberation and interpersonal contexts, such as being at least somewhat motivated to do what one believes one ought to do, being disposed to blame others for actions one takes to be wrong, etc.

Someone might object that this alleged connection between concept identity and normative role is too tight, because different people or communities can use the same normative concepts while exhibiting slightly different behavioral and affective tendencies. For instance, two communities might use the same concept of wrongness even though, in one case, judgments of wrongness are strongly associated with a disposition to feel and express anger at perceived wrongdoers, while for the other community this connection between judgments of wrongness and anger is weaker or even absent.

Strictly speaking, this worry is not relevant to the metasemantic challenge as we have understood it so far: quasi-naturalists can still account for Conceptual Elasticity, Referential Stability, and Objectivity in cases like Moral Twin Earth and Honor Code, where by stipulation the relevant concepts *do* have the same normative roles for all communities involved. To put it differently, the quasi-naturalist response to the metasemantic challenge only requires the claim that normative concepts are identical *if* they have the same normative roles, and the objection stated above does not affect this claim.

However, this response would only push the bump under the rug, because the worry can persist by reframing the very setup of the challenge: we can imagine that even in scenarios like Moral Twin Earth and Honor Code, the given communities exhibit similar but slightly different motivational patterns in connection with the use of the relevant normative concepts, and quasi-naturalists arguably still face the challenge of making good on Conceptual Elasticity, Referential Stability and Objectivity in such cases.²⁸

Quasi-naturalists might be tempted to address this problem by loosening the identity conditions for normative concepts. That is, they might argue that different communities can use the same normative concepts even if they assign slightly different normative roles to those concepts, because sameness of concepts does not require sameness of normative roles but rather *enough similarity* between normative roles. This suggestion seems compatible with expressivism: for instance, expressivists might claim that the concept of *ought* encodes the same type of desire-like mental state in different communities, such as an attitude of disapproval of actions that have certain natural properties, even if this mental state manifests itself in slightly different affective and behavioral tendencies in

interpret claims about normative reference and objectivity as first-order normative claims amenable to an expressivist account. I believe hybrid expressivism *is* needed for a successful quasi-naturalist response to the metasemantic challenge, but only because it plays a key role in accounting for the metaphysical claims of normative naturalism in the first place, as I pointed out above.

²⁸Eklund (2017, pp. 56-57) raises a similar worry for views according to which concepts with the same normative role thereby have the same reference, which he calls “the embarrassment of riches problem”.

those communities.

However, this idea that sameness of normative concepts only requires sufficient similarity between normative roles seems vulnerable to a sorites: we can imagine a sequence of normative concepts such that each two adjacent concepts have sufficiently similar normative roles to count as the same concept but the first and the last concept in the series have very different normative roles. Given that identity is transitive, it would follow that these concepts with very different normative roles are nevertheless the same concept, which should be unacceptable to expressivists.²⁹

Perhaps quasi-naturalists can address this issue: for instance, they might propose that sameness of normative concepts requires not only enough similarity in normative roles, but also that those similar normative roles include a shared set of core affective and behavioral tendencies. But there is a better option for dealing with the original worry about normative concepts with slightly different normative roles, which avoids this sorites threat. Quasi-naturalists can claim that normative concepts do indeed differ whenever their normative roles differ, but if those differences are minor something in the vicinity of Conceptual Elasticity still holds: the relevant communities are thinking and talking about *a common normative subject-matter* even if they are using slightly different concepts, because those concepts govern their practical lives in similar enough ways. After all, what matters from an expressivist standpoint is not sameness of concepts as such, but rather the possibility of normative disagreement understood as actual or potential practical conflict between desire-like mental states, and concepts with similar enough normative roles do allow for such practical conflict. Once genuine disagreement is secured in these cases where different communities use concepts with slightly different normative roles, quasi-naturalists can again address any questions about the reference of those concepts by engaging in first-order normative theory, which may well deliver unique and objective answers.

A different objection to this quasi-naturalist proposal might focus on its deflationary framework, which seems to dodge or dissolve important theoretical questions about normative reference. For instance, what is it in virtue of which normative terms share a reference in cases like Moral Twin Earth or Honor Code? Quasi-naturalists have not offered any substantive answer to this question, the objection would go, and more generally no illuminating characterization of the relation of normative reference.

Now, deflationism rejects the idea of reference as a metaphysical relation between terms/concepts and the world which might play a substantive explanatory role in the theory of meaning: on this view, the notion of reference, just like the notion of truth, is merely an expressive device that allows us to endorse and reject first-order claims in the given domain of discourse, and to abbreviate and generalize when talking about claims that we endorse or reject. So it is true that deflationists do not have an interesting story to tell about the nature of normative reference as such (though of course they do have interesting things to say about the *notion* of reference and its function). However, deflationism does allow for substantive answers to any *specific* questions about normative reference, like the question of what makes it the case that “wrong” as used on Moral Twin

²⁹Thanks to an anonymous reviewer for raising this worry.

Earth refers to the same natural property as our “wrong,” namely *normative* answers, because those questions themselves are interpreted as first-order normative questions in a deflationary framework. The thesis that the same natural features make actions wrong for us and on Moral Twin Earth, for instance, is not a claim that somehow trivially follows from deflationism; arguing for it requires substantive work in normative theory.³⁰

Finally, someone might complain that the quasi-naturalists’ approach to questions about normative reference entails that people can be vastly ignorant about their normative terms refer to, insofar as they are vastly mistaken in their normative beliefs. Moreover, this consequence of quasi-naturalism goes against an epistemic access condition on reference that seems plausible in virtually all other domains of discourse.³¹

The short answer to this objection (the only one I can give here) is that this is indeed what quasi-naturalism holds: normative reference works differently than reference in other areas of discourse, and in particular people can be in the dark about what their normative terms refer to, in contrast to ordinary descriptive language, where an epistemic access condition on reference is exceedingly plausible. It is a core idea of the expressivist project that the primary function of normative discourse is practical rather than representational, and this is why different people can use the same normative concepts even while applying them to very different things in the world, as long as those concepts play roughly the same practical function in their lives. If we combine this core expressivist idea with a deflationary picture on which questions about truth and reference are treated exclusively as internal to normative theorizing, it should be no surprise that we end up with a view on which the reference of someone’s normative terms can be radically at odds with their normative judgments and their beliefs about what those terms refer to.

I have little hope that the last two responses will allay anyone’s worries about expressivism, deflationism and their upshot for the metasemantics of normative discourse. But let me end by pointing out that it is precisely these features of quasi-naturalism—the rejection of a use-dependent conception of reference, and the deflationary interpretation of all questions about normative reference as internal to normative discourse—that allow for a full solution to the metasemantic challenge to normative naturalism.³²

³⁰It might be useful to note that normative theorizing need not always lead to objectivist conclusions or verdicts of sameness of reference in cases where different communities vastly diverge in their uses of normative concepts: there might be good reasons to adopt some form of relativism with respect to some such cases, like a view according to which the reference of “wrong” has shifted across human history due to radical changes in the circumstances and needs of human beings. Whether such a relativism of distance is ultimately plausible is again a matter to be settled through normative theory. The quasi-naturalist response to the metasemantic challenge only involves the claim that objectivism about reference *can* be the right stance with respect to cases like Moral Twin Earth and Honor Code.

³¹See Schroeter and Schroeter (2013) for an in-depth discussion of the epistemic access condition on reference and a related argument to the effect that an account of normative reference should align with a general, use-dependent conception of reference applicable to all areas of discourse.

³²For helpful comments and discussion on previous versions of this paper, many thanks to Farbod Akhlaghi, James Brown, Claire Kirwin, Manuel García-Carpintero, Will Gamester, Nick Laskowski, Teresa Marques, Sven Rosenkranz, Alex Sandgren, Thomas Schmidt, Pekka Väyrynen, Robbie Williams, Jack Woods, audiences at the University of Leeds, the University of Barcelona, the 2021 Cyprus Metaethics Workshop, and the 2022 “Truth in Evaluation” Conference of the Italian Society for Analytic Philosophy, and two anonymous reviewers for *Erkenntnis*. Research for this paper was funded through the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 837036.

References

1. Artiga, Marc. 2015. "Rescuing Tracking Theories of Morality." *Philosophical Studies* 172: 3357-3374.
2. Bex-Priestley, Graham. Forthcoming. "Expressivists Should Be Reductive Naturalists." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 17. Oxford University Press.
3. Blackburn, Simon. 1993. *Essays in Quasi-Realism*. Oxford University Press.
4. Boyd, Richard. 1988. "How to Be a Moral Realist." In G. Sayre-McCord (ed.), *Essays on Moral Realism*, Cornell University Press, 181-228.
5. Boyd, Richard. 1993. "Metaphor and Theory Change." In A. Ortony (ed.), *Metaphor and Thought*, 2nd Edition. Cambridge University Press.
6. Boyd, Richard. 2003. "Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism. Part I." *Philosophy and Phenomenological Research* 66 (3): 505-553.
7. Brink, David O. 2001. "Realism, Naturalism, and Moral Semantics." *Social Philosophy and Policy* 18: 154-176.
8. Copp, David. 2000. "Milk, Honey, and the Good Life on Moral Twin Earth." *Synthese* 124 (1-2): 113-137.
9. Copp, David. 2001. "Realist-Expressivism: A Neglected Option for Moral Realism." *Social Philosophy and Policy* 18: 1-43.
10. Copp, David. 2008. "Darwinian Skepticism About Moral Realism." *Philosophical Issues* 18: 186-206.
11. Copp, David. 2018. "Realist-Expressivism and the Fundamental Role of Normative Belief." *Philosophical Studies* 175: 1333-1356.
12. Curry, Oliver Scott. 2016. "Morality as Cooperation: A Problem-Centred Approach." In *The Evolution of Morality*, 27-51. Springer.
13. Dowell, Janice. 2016. "The Metaethical Insignificance of Moral Twin Earth." In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 11, Oxford University Press.
14. Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford University Press.
15. Finlay, Stephen. 2014. *Confusion of Tongues: A Theory of Normative Language*. Oxford University Press.
16. Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.
17. Gibbard, Allan. 2003. *Thinking How to Live*. Harvard University Press.

18. Hare, Richard M. 1952. *The Language of Morals*. Oxford Clarendon Press.
19. Horgan, Terence and Timmons, Mark. 1991. "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* 16: 447-465.
20. Horgan, Terence and Timmons, Mark. 2000. "Copping Out on Moral Twin Earth." *Synthese* 124 (1-2): 139-152.
21. Kripke, Saul A. 1972. *Naming and Necessity*. In *Semantics of Natural Language*, Springer, Dordrecht, 253-355.
22. Laskowski, N. G. 2018. "Reductivism, Nonreductivism and Incredulity About Streumer's Error Theory." *Analysis* 78 (4): 766-776.
23. Laskowski, N. G. 2019. "The Sense of Incredibility in Ethics." *Philosophical Studies* 176 (1): 93-115.
24. Laskowski, N. G. 2020. "Resisting Reductive Naturalism." In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 15, Oxford University Press.
25. Locke, Dustin. 2014. "Darwinian Normative Skepticism." In Michael Bergmann and Patrick Kain (eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution*. Oxford University Press.
26. MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford University Press.
27. McPherson, Tristram. 2013. "Semantic Challenges to Normative Realism." *Philosophy Compass* 8 (2): 126-136.
28. Merli, David. 2002. "Return to Moral Twin Earth." *Canadian Journal of Philosophy* 32 (2): 207-240.
29. Morton, Justin. 2018. "When Do Replies to the Evolutionary Debunking Argument Against Moral Realism Beg the Question?" *Australasian Journal of Philosophy* 97: 265-280.
30. Parfit, Derek. 2011. *On What Matters*, vol. 2. Oxford University Press.
31. Putnam, Hilary. 1975. "The Meaning of 'Meaning'." *Minnesota Studies in the Philosophy of Science* 7: 131-193.
32. Railton, Peter. 2017. "Two Sides of the Meta-ethical Mountain?" In Peter Singer (ed.), *Does Anything Really Matter? Essays on Parfit on Objectivity*, Oxford University Press,
33. Ridge, Michael. 2007. "Ecumenical Expressivism: The Best of Both Worlds?" In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 2, 51-76.
34. Ridge, Michael. 2014. *Impassioned Belief*. Oxford University Press.
35. Sayre-McCord, G. 1997. "'Good' on Twin Earth." *Philosophical Issues* 8: 267-292.

36. Schroeder, Mark. 2013. "Tempered Expressivism." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 8. Oxford University Press.
37. Schroeder, Mark. 2014. "The Truth in Hybrid Semantics." In Guy Fletcher and Michael Ridge (eds.), *Having it Both Ways: Hybrid Theories and Modern Metaethics*, Oxford University Press, 273-293.
38. Schroeter, Laura and Schroeter, François. 2013. "Normative Realism: Co-reference without Convergence?" *Philosophers' Imprint* 13 (13): 1-24.
39. Sterelny, Kim and Fraser, Ben. 2016. "Evolution and Moral Realism." *British Journal for the Philosophy of Science* 68: 981-1006.
40. Streumer, Bart. 2017. *Unbelievable Errors: An Error Theory About All Normative Judgments*. Oxford University Press.
41. Toppinen, Teemu. 2013. "Believing in Expressivism." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 8, Oxford University Press.
42. van Roojen, Mark. 2006. "Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 1, 161-194.
43. Väyrynen, Pekka. 2018a. "A Simple Escape From Moral Twin Earth." *Thought* 7: 109–118.
44. Väyrynen, Pekka. 2018b. "Normative Commitments in Metanormative Theory. In Jussi Suikkanen & Antti Kauppinen (eds.), *Methodology and Moral Philosophy*, Routledge, 193-213.
45. Williams, J. Robert G. 2018. "Normative Reference Magnets." *Philosophical Review* 127 (1): 41-71.
46. Williams, J. Robert G. 2020. *The Metaphysics of Representation*. Oxford University Press.
47. Wong, David B. 2006. *Natural Moralities: A Defense of Pluralistic Relativism*. Oxford University Press.