

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Schroeter, L;Schroeter, F

Title:

The Generalized Integration Challenge in Metaethics

Date:

2019-03-01

Citation:

Schroeter, L. & Schroeter, F. (2019). The Generalized Integration Challenge in Metaethics. *Nous*, 53 (1), pp.192-223. <https://doi.org/10.1111/nous.12216>.

Persistent Link:

<https://hdl.handle.net/11343/293314>

The Generalized Integration Challenge in Metaethics

Laura Schroeter and François Schroeter

The Generalized Integration Challenge

The University of Melbourne

Abstract: The Generalized Integration Challenge (GIC) is the task of providing, for a given domain of discourse, a simultaneously acceptable metaphysics, epistemology and metasemantics and showing them to be so. In this paper, we focus on a metaethical position for which (GIC) seems particularly acute: the brand of normative realism which takes normative properties to be (i) mind-independent and (ii) causally inert. The problem is that these metaphysical commitments seem to make normative knowledge impossible. We suggest that bringing metasemantics into play can help to resolve this puzzle. We propose an independently plausible metasemantic constraint on reference determination and show how it can provide a plausible response to (GIC) for this brand of normative realism.

The concept of mathematical truth, as explicated, must fit into an over-all account of knowledge in a way that makes it intelligible how we have the mathematical knowledge that we have. (Benacerraf 1973, 667)

I call the general task of providing, for a given area, a simultaneously acceptable metaphysics and epistemology, and showing them to be so, the Integration Challenge for that area. (Peacocke 1999, 1)

Imagine a realm of objects and properties that are both mind-independent and causally inert. Could we have knowledge of this domain? On the face of it, it seems not. The picture that naturally arises is of two hermetically sealed domains, with no possibility of information transfer between the two. If we're trapped on one side of this barrier while the independent objects and properties are arrayed on the other side, it seems impossible for us to have non-accidentally true beliefs about them.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/nous.12216](#).

This article is protected by copyright. All rights reserved.

This type problem is familiar territory for Platonists in the philosophy of mathematics (Benacerraf 1973; Field 1989). As Paul Benacerraf notes,

[the Platonist lacks] an account of the link between our cognitive faculties and the objects known. [...] [S]omething must be said to bridge the chasm [...] between the entities that form the subject matter of mathematics and the human knower [...]" (1973, 674-5).

David Enoch (2010) has pointed out that a similar problem arises in metaethics for non-naturalist normative realists who take normative properties to be mind-independent and causally inert: it seems utterly mysterious how our normative beliefs could be non-accidentally true.

In this paper, we'll follow Christopher Peacocke's general characterization of the problem highlighted by Benacerraf. According to Peacocke, the Benacerraf problem is just a special case of what he calls *the Integration Challenge*: the epistemological theory and the metaphysical theory for a domain must be integrated in a plausible way to explain how we can know statements in that domain, when we know them (Peacocke 1999, 1). Since Platonists are committed to a specific metaphysics of mathematical objects, they must show how our actual belief-forming dispositions could yield knowledge of that sort of thing. This challenge of explaining the possibility of a knowledge relation between a belief and the objects and properties that belief represents faces realists in *any* domain.

We suggest that a third theoretical element must be integrated with the others, metasemantics. Our focus here is on the level of idiolect meaning and conceptual content. At this level, a metasemantic theory explains why the attitudes expressed by particular token uses of expressions have the semantic contents that they do. For instance, a metasemantic theory might explain why your token use of the expression 'Socrates' on a particular occasion picks out a specific individual, Socrates, and why your token use of the expression 'is wise' attributes a particular property, wisdom. Metasemantics thus provides an explanation of the fact that we talk and think about certain types of objects and properties. Explaining this aboutness relation between an attitude and a particular object or property can be just as problematic as explaining the knowledge relation – or more so. Alternatively, the aboutness relation may prove more theoretically tractable, and provide further resources for resolving the Integration Challenge. That's the possibility we will be exploring in this paper.

The role of metasemantics in the Integration Challenge hasn't gone unrecognized. Indeed, Peacocke himself recruits his own broadly Fregean metasemantic framework to explain knowledge in specific domains. When our beliefs involve 'epistemically-individuated' concepts,

Peacocke argues, the relevant belief-forming dispositions will be reliable (Peacocke 1999, chap. 2). But what we'd like to stress is that taking metasemantics into account is not just one potential strategy for integrating metaphysics and epistemology within certain domains. A metasemantic theory – whether Fregean or causal or interpretivist or deflationary – is an essential ingredient in any adequate philosophical account of a domain. A comprehensive philosophical account of mathematics or morals must explain how the metaphysics, epistemology and metasemantics are integrated into a coherent account of the nature of the target domain and our epistemic and semantic access to facts in that domain. We call this task the *Generalized Integration Challenge*.

In this paper we concentrate exclusively on the metaethical case. The first section introduces the Generalized Integration Challenge (GIC) and clarifies the type of normative realism we take as our test case. In section two, we highlight a metasemantic commitment, Minimal Charity, that's widely shared in philosophy of mind and language. We show how this commitment places constraints on the scope for error and thus constitute a first step towards addressing the Integration Challenge in metaethics. However, we show that Minimal Charity falls well short of resolving (GIC) for realism in any domain. In Section three, we introduce a significantly strengthened version of Charity, Ideal Accessibility, and explain how this constraint can help meet (GIC). But the devil is in the details. The next two sections take a closer look at the epistemological and metaphysical sides of (GIC) in our test case. We examine whether this strong metasemantic constraint can vindicate the specific epistemic and metaphysical commitments distinctive of the normative realist position we have chosen as our target. After reviewing our argument in section 6, we close in section 7 by considering whether the metasemantic strategy for answering (GIC) can be extended to a different type of metasemantic theory.

1. The Generalized Integration Challenge

Peacocke characterizes the Integration Challenge as follows:

(IC) The task of providing, for a given area, a simultaneously acceptable metaphysics and epistemology, and showing them to be so. (1999, 1)

As Peacocke notes, (IC) does not assume that we *do* have knowledge about a domain – we may not. (IC) simply requires that the epistemological theory and the metaphysical theory for a domain be integrated in a plausible way. (IC) is also neutral between realist and anti-realist metaphysics. Even if one rejects the existence of normative properties that make normative judgments true, one must still explain the metaphysics and epistemology of normative judgment: what's the nature of these judgments, which features of the world are our actual judgment-forming dispositions sensitive to, and what is the epistemological status of those judgments? If there is no plausible integrated answer to these questions, we should favor *ceteris paribus* alternative theories that can better satisfy (IC).

The Generalized Integration Challenge adds metasemantics to the list of philosophical theories that must be integrated in a plausible way:

(GIC) The task of providing, for a given area, a simultaneously acceptable metaphysics, epistemology and metasemantics, and showing them to be so.

Here again, the challenge is meant to be general: it does not assume that we have knowledge about a domain, and it is neutral between different metaphysical or metasemantic accounts of the target domain. The rationale for the addition of metasemantics to (IC) is simple.

Metasemantics explains the link between belief states and their contents – the two basic relations of the knowledge relation. Normative realists, for instance, posit normative properties and assume that normative terms such as ‘is right’ pick out these properties. They thus owe us a metasemantic account of how reference to normative properties is secured. Theories of reference determination have their own constraints, which are distinct from those of epistemology or metaphysics. So a referentialist semantics for normative terms must fit within a plausible overall metasemantic theory of how different sorts of terms acquire their semantic contents.ⁱ

Metasemantics is sometimes understood more narrowly than we do here. Seth Yalcin, for instance, reserves the term ‘semantics’ for a formal compositional semantics for a particular language (Yalcin 2014). The goal of compositional semantics is to model how individual expression types interact and affect the meaning of whole sentences. But compositional semantics does not care which objects or properties are actually picked out by a given expression – or indeed if any such objects or properties are picked out. Given this narrow understanding of ‘semantics’, Yalcin takes ‘metasemantics’ to be a theory that explains how an individual’s psychology – her language faculty – determines that particular expressions in her lexicon play specific formal semantic roles. Thus, on Yalcin’s approach, semantics and metasemantics prescind from any claims about the substantive contents words express or communicate. Some normative expressivists have recently argued that they can accept the same formal compositional semantics as realists: so formal semantics and narrow metasemantics are strictly neutral about whether a subject’s use of ‘this is right’ attributes a specific property or expresses a non-cognitive attitude toward a salient object (Ridge 2014; Chrisman 2015; Pérez Carballo 2015; Silk 2015).

We understand ‘semantic content’ and ‘metasemantics’ in a broader sense, which focuses primarily on the level of thought rather than language. The substantive semantic content of particular expressions in an individual’s idiolect depends on the content of stably associated concepts. So a metasemantic theory in the broad sense should explain why the *concepts* expressed by token uses of expressions have the substantive semantic contents that they do.ⁱⁱ On a referentialist semantics for normative terms, for instance, the semantic content of ‘is right’ will be a specific property with specific empirical instantiation conditions. The metasemantic

theory must explain how aspects of the subject's understanding and history of use (perhaps together with facts about her actual social, psychological, or physical environment) suffice to single out *this specific property* among all possible properties as the reference of that term. For a contextualist or a relativist semantics, the metasemantic theory must explain how a subject's understanding and use suffice to single out *this specific function* from empirical facts about the context of use and/or the context of assessment to a particular property as the semantic content standardly expressed by 'is right'. And for an expressivist semantics, the metasemantics must explain which aspects of understanding and use suffice to single out *this very attitude* among all possible attitudes as the thought content standardly expressed by 'is right'. It's common ground among all metaethical theorists that expressions like 'is right' have a stable, shared semantic content in this sense – that normative terms make a stable contribution to the thought contents expressed by their use on particular occasions.ⁱⁱⁱ We take the job of metasemantics to be that of explaining how this stable substantive semantic content is determined.

As we suggested, (GIC) applies to all metaethical positions. In this paper, however, we'll focus on one metaethical position for which (GIC) has been thought to be particularly acute: normative realism that takes normative terms to pick out properties that are both *mind-independent* and *causally inert*. These are the metaphysical features that gave rise to Benacerraf's epistemic challenge to mathematical Platonism. If the objects of knowledge are wholly independent of our psychological states and cannot causally interact with psychological states, then it seems mysterious how we could come to know about them. Hartry Field famously dramatized the problem with an analogy: it's as if we had true beliefs about events in a remote Nepalese village. If there is no information flow from the village to us, it's utterly mysterious how our beliefs could qualify as knowledge. Our knowledge of mind-independent causally inert mathematical and moral facts seems similarly mysterious.

It is a delicate matter how exactly to characterize mind-independence and causal inertia – we'll return to these issues below in section 5. We take mind-independence and causal inertness to be metaphysical features of properties. To a first approximation, a property is mind-independent iff its instantiation by an object does not counterfactually depend on the instantiation of characteristic psychological states like judgments, motivations, or emotions about the object. For instance, the property of *being right* is mind-independent if there are possible worlds in which certain actions are morally right, but the characteristic psychological states associated with taking an action to be right are absent: e.g. no one in that world judges those actions right, feels approval of them, or is disposed to promote them. A property is *causally inert* iff it does not figure essentially in our best causal explanations of events. For instance, if we do not need to cite the fact that an action is right when explaining physical events (like why the action caused the dam to open) or mental events (like the judgment that it is right), the property of *being right* is causally inert.

Many normative realists take normative terms to pick out properties that are mind-independent and causally inert in this sense. We'll call the family of realist positions that accept these theses *Autonomous Normative Realism* (ANR). Non-naturalists are paradigm proponents of (ANR). Non-naturalists include both Robust Realists like (FitzPatrick 2008; Enoch 2011) who take metaphysics to play a crucial role in explaining our fallibility about normative facts, and Anti-Metaphysical Realists like (Parfit 2011; Scanlon 2014) who take our normative practices to determine the nature of normative properties. However, both types of non-naturalism agree that the instantiation of normative properties is independent of specific psychological states and that normative properties don't play an important causal-explanatory role.^{iv} Naturalists can also accept (ANR). One can be a naturalist about restaurants, for instance, without supposing that *being a restaurant* is a mind-dependent property or that it plays an essential role in our causal-explanatory theory of the world (Clarke-Doane 2012, 318). In this paper, we will present a metasemantic position that yields just this type of naturalist realist view of normative properties.

In what follows, we'll treat (ANR) as a fixed point, and work towards an explanation of knowledge of the normative domain. Our question is whether metasemantics can help bridge the 'chasm' Benacerraf highlighted between the objects and properties characteristic of a given domain and our beliefs about them. A full epistemic response to (GIC) would require an explanation of topics like normative knowledge, justification, and epistemic responsibility. In this paper, however, we'll restrict our attention to explaining the *reliability* of our normative beliefs – or more precisely of the processes and dispositions generating them. This restriction will make our discussion more tractable, since we can set aside controversies about how to understand contested epistemological properties. And it brings our discussion into line with the standard debate in the philosophy of mathematics that has been generalized to metaethics (e.g. Field 1989; Enoch 2010; Clarke-Doane 2012, 2014). Most importantly, reliability provides a good benchmark for progress in bridging the chasm highlighted by Benacerraf. The proponent of (ANR), after all, holds that our beliefs about mind-independent causally inert normative properties are often true, and an explanation of the reliability of our beliefs would go a long way towards dispelling the epistemological mystery of how this could be the case.^v In contrast, focusing on epistemological notions like knowledge or justification, rather than reliability, can lead one to appeal to notions like self-evidence, intuition, testimony, or coherence. Such approaches fail to respond to (GIC): they cannot establish a bridge between psychology and metaphysics, since they operate entirely on the psychological side of Benacerraf's chasm.^{vi}

What if no explanation of the reliability of our mathematical (or normative) beliefs is possible in the case of mind-independent causally inert properties? According to Field, skepticism would follow: (i) beliefs about such properties would be unjustified and (ii) recognition of this fact would tend to erode our psychological commitment to such beliefs. As many authors have pointed out, these strong conclusions do not follow: beliefs may be justified even if we can't explain their reliability, and we may psychologically hold onto such beliefs even if we know we cannot explain their reliability (for relevant discussion, see Enoch 2010; Setiya 2012, chap. 2;

Clarke-Doane 2015; Tersman 2015). But the force of Field's Nepalese village analogy does not depend on these overly strong conclusions. At bottom, the challenge is to integrate an account of our epistemic practices with an account of the metaphysics of the domain in question and our ability to talk and think about it. Failure to answer the General Integration Challenge comes at a significant *theoretical cost* to a theory's overall plausibility. If mathematical Platonism cannot explain the reliability of our dispositions to form mathematical beliefs, it is less plausible *ceteris paribus* than rival accounts that can explain this reliability. So Field's conclusion should simply be that mathematical Platonism provides a less plausible answer to (GIC) than alternative views (such as his own anti-realism). In general, any philosophical position that involves a brute mystery in how its metaphysics, epistemology, or metasemantics are integrated is to that extent a less attractive theory.

2. Bridging the chasm: aboutness relations

The image of a metaphysically independent, hermetically sealed domain of normative properties makes it seem like a miracle that any of our beliefs about that domain are true. But it is important to see that the two sides of this chasm – the psychological states and the normative properties – are linked by a semantic *aboutness* relation. When we consider the reliability of normative beliefs, we are focusing on a very specific class of psychological states: beliefs that involve normative concepts that pick out particular properties.^{vii} What secures this aboutness relation?

To begin answering this question, consider the claim that the concept you express with the predicate 'is morally right' picks out the property of *being an even number*. This interpretation is absurd and no metasemantic theory of reference determination would endorse it. Why?

The obvious response is that the proposed interpretation has nothing to do with your actual understanding and use of the expression 'is morally right': you never concatenate this expression with a singular term denoting a number, your reasoning about whether 'is morally right' is true doesn't turn on pure mathematical reasoning, the testimony you take to be authoritative in deciding the truth of such sentences often involves an assessment of the authority's character rather than her logical acumen, and so on. In short, the proposed interpretation would construe your moral beliefs and associated cognitive dispositions as systematically mistaken in every way. Moreover, the observation generalizes. It would be similarly crazy to interpret your use of the term 'is a glacier' as picking out the property of *being an electron*.

This observation should give us pause. It suggests a minimal constraint on a referential interpretation of a particular concept:

Minimal Charity: The semantic interpretation of a concept must make at least *some* elements of the subject's total understanding of that concept come out true or truth-preserving.

Minimal Charity imposes the very weak requirement that at least *some* of the states that constitute a subject's total understanding of the reference – e.g. her beliefs, intuitions, inferential dispositions, or reflective dispositions – must be accurate.^{viii} This requirement is significantly weaker than Quine's 'Principle of Charity', which requires the interpreter to maximize the truth of the totality of sentences the subject would assent to (Quine 1960). In particular, Minimal Charity takes no stand on the proportion of true to false beliefs: your true beliefs about electrons could be very few and far between. This means Minimal Charity would be woefully inadequate as a full theory of reference-determination. It merely imposes a weak necessary condition for referential assignments: you can't be wholly off track in every respect. Unlike Quine's thesis, moreover, Minimal Charity can take into account non-doxastic states like inferential dispositions or intellectual seemings, whose content may not be easily formulated in natural language and which the subject may not be immediately disposed to assent to if they were.

We've motivated Minimal Charity by appealing to examples, but it's important to see that it is a theoretically motivated constraint. To see why, it's helpful to consider the rationale for characterizing a subject's thought and talk in terms of its semantic content. Theorists working on mental content agree that content ascriptions play two central theoretical roles:

- (i) Normative assessment: Ascribing specific semantic contents to mental states allows us to assess the truth of a subject's beliefs, the success of her intentions, and the rationality of her reasoning.
- (ii) Prediction and explanation: Ascribing specific semantic contents to mental states puts us in a position to predict and explain a subject's behavior and reasoning.

To play these roles, content ascriptions must reflect *some* aspect of the subject's dispositions: the subject's implicit understanding of what it is she's thinking about cannot be entirely off-track. If we assume that the correct semantic interpretation of the beliefs you express with the term 'water' needn't construe any of your cognitive states as accurate, it's hard to see how that interpretation could help to predict and explain your actions. The more accurate we assume your understanding of 'water' is, in fact, the more helpful this interpretation will be in predicting and explaining your behavior. Likewise, it's hard to see how a content ascription that has nothing to do with your own understanding and use of the term 'water' could provide appropriately authoritative standards for criticizing your beliefs or reasoning methods as mistaken. Why should you listen to someone's critique of your claim, 'water is essential to survival', if they're interpreting the belief you express as affirming the proposition that long division is essential to badminton?

Contemporary metasemantic theories, almost without exception, respect Minimal Charity. It's helpful to divide them according to which of the two central theoretical roles they privilege. Those who focus on the first role, normative assessment, impose much more demanding charity constraints on semantic interpretation, tying referential assignments to the vindication of the subject's current understanding.^{ix} The second class of theories treats scientific prediction and explanation of behavior as the primary constraint on the determination of semantic contents. Theorists in this camp tend to favor naturalistic reductions of the reference relation. Information theoretic accounts (Stampe 1977; Dretske 1981; Fodor 1987), for instance, ground semantic interpretation in reliable lawlike correlations between subjects' recognitional judgments ('there's some water!') and the occurrence of certain properties in their environment (H₂O at the indicated spot). Such accounts guarantee that at least some of the subject's recognitional dispositions will be accurate, and that any judgments formed in this way will be true.^x

Minimal Charity makes the chasm between normative beliefs and normative properties look less daunting. It establishes a necessary connection between representational states and the states of affairs they represent: it's metaphysically necessary that some of the attitudes and dispositions associated with entertaining a concept be true or truth-preserving with respect to the object, kind, relation, truth function, or other feature represented by that concept. This metasemantic constraint guarantees a minimal level of accuracy in the subject's overall understanding of the reference. This internal connection between the two sides of Benacerraf's chasm suggests that metasemantics may provide a way forward for proponents of (ANR) to answer (GIC).^{xi}

However, Minimal Charity falls well short of establishing that we have reliably true beliefs about which actions are right and wrong. Even if some elements of the subject's total understanding must be accurate, it clearly doesn't follow that her beliefs in the target domain will be mostly true (or true relevantly more often than not). Moreover, in this section we have simply assumed that Autonomous Normative Realism is true, and thus that normative concepts do semantically pick out particular properties. *If* we are able to represent mind-independent, causally inert normative properties, we've argued, *then* Minimal Charity ensures that some of our normative beliefs or cognitive dispositions are accurate. But why suppose that the antecedent of this conditional can ever be satisfied – that our psychological states can pick out properties on the other side of Benacerraf's chasm? One may worry that the original explanatory challenge will simply reemerge at the level of reference determination. The metasemantic story must go significantly beyond Minimal Charity if it's to assuage the worry that determinate reference is impossible without causal or constitutive relations linking our psychological states to the specific objects, kinds, or properties they allegedly represent.

3. Strengthening Minimal Charity

To answer these challenges, we'll need to abandon our metasemantic neutrality. We'll focus on an influential metasemantic constraint that we find plausible and show how it can explain reference to mind-independent, causally inert normative properties. In the following section, we'll then examine whether this metasemantic constraint can help establish the reliability of our beliefs about which actions are right and wrong.

A central goal of metasemantic theorists who privilege the role of rational assessment is to vindicate the subject's epistemic commitments about what her words and thoughts represent. One way to capture these commitments is to require referential assignments to respect the subject's own ideal, fully informed verdicts about the reference of her words and concepts. Consider the following Ideal Accessibility constraint:

Ideal Accessibility (IA): The correct semantic interpretation of the referential concept expressed by a term 'X' must make the subject's ideal, empirically-informed beliefs about which things are X in any possible world come out true.

(IA) is based on a proposal in (Chalmers and Jackson 2001). But the constraint is characteristic of traditional metasemantic theories that privilege the normative role of content in assessing beliefs as true or false and in setting a standard for epistemic success.^{xii} (IA) targets the subject's *use* of the concept expressed by 'X' to identify the property's instantiations in various possible worlds. To avoid circularity, the empirical information about these worlds must be couched in a more basic descriptive vocabulary that excludes the ordinary language term in question. For instance, Chalmers and Jackson take this more basic descriptive vocabulary to be an exhaustive microphysical description of the actual world, together with a 'that's all' clause and an 'I am here' clause.^{xiii} The idea is that such a basic description would fix all the empirical facts that are relevant to determining the reference of your ordinary words and thoughts. Ideal Accessibility then says that the reference of a subject's non-base-level concepts is determined by the subject's own fully informed, ideally reflective verdicts about the reference.

For example, consider the concept expressed by your term 'water'. Full information about your actual empirical circumstances would locate you with respect to your physical, social, and historical environment described in the minutest physical detail. This physical description contains all the information needed to identify macro-level physical facts, such as the fact that the stuff you have been classifying as 'water' has been largely composed of H₂O and that aggregations of H₂O display the macro-level properties you associate with 'water' (clear, potable, falls as rain, etc.). Suppose this macro-level information, together with your prior understanding of 'water' would ideally lead you to conclude that water = H₂O. According to Ideal Accessibility, then, this verdict is correct: the correct semantic interpretation of your words and thoughts can't clash with ideally reflective verdicts based on your ordinary understanding.^{xiv}

We mention Chalmers's appeal to micro-level descriptions for illustrative purposes only. In the case of normative terms like 'is right', the base vocabulary sufficient for deciding which things are right would not need to descend to the level of microphysics. Any descriptive vocabulary rich enough to generate a complete characterization of the supervenience base of normative properties would do.

Ideal Accessibility is a strengthening of Minimal Charity. This may not be obvious, since Minimal Charity constrains the subject's actual beliefs and cognitive dispositions, whereas Ideal Accessibility constrains the subject's ideal beliefs. But ideal beliefs about X must be justifiable on the basis of the subject's current understanding of 'X' on pain of changing the topic. So Ideal Accessibility constrains the subject's actual understanding of 'X' indirectly: all of the subject's actual attitudes and dispositions that would ultimately figure in justifying her ideal, fully-informed verdicts about which things are X must be roughly true or truth-preserving. This is a stronger constraint than the minimal requirement that *some* beliefs and inferential dispositions come out roughly true or truth-preserving. Moreover, this requirement prioritizes among the subject's actual beliefs, selecting those that would survive ideally informed reflection as the accurate ones.

To get a feeling for how Ideal Accessibility applies in the normative case, consider how you would ideally come to verdicts about which actions count as right. Your starting point will be your total current understanding of the term, which includes your implicit opinions about which considerations are relevant to determining the normative status of an action, as well as your dispositions to correct those opinions in the light of empirical discoveries, others' criticisms and testimony, the genealogy of your use of the term, and systematic reflection. This starting point will push you towards a reflective refinement of your understanding. Ideally, this process will lead to a point of reflective equilibrium, in which your dispositions to identify actions as right in various circumstances will not be overturned by further information or reflection. At this point of equilibrium, we can use your dispositions to fix an extension for 'right' in every possible world. For any empirical circumstance, you will be disposed to judge a given action right, or you'll be unable to come to a determinate verdict. These ideal verdicts fix an extension in all possible worlds (an intension) for your use of the predicate 'is right' (in cases where you cannot come to a settled verdict about particular cases, the extension will be indeterminate with respect to those cases). According to Ideal Accessibility, these verdicts are guaranteed to be true.

The key point for present purposes is that Ideal Accessibility can help explain reference to mind-independent causally inert normative properties posited by (ANR). (IA) identifies a unique property N as picked out by your predicate 'N' given the following four assumptions:

1. Determinacy: The subject's understanding of 'N' is sufficiently disciplined to generate a point of ideal reflective equilibrium about the instantiation of N in all possible worlds.

2. Global supervenience: In ideal reflective equilibrium, the subject takes the instantiation of N to be fully determined by more basic descriptive facts about a situation.
3. Plenitudinous properties: A set-theoretic construct specifying an extension for a predicate in all possible worlds picks out a property.
4. Property identity: Properties are identical iff they have exactly the same instances in all possible worlds.

Conditions (1) and (2) place constraints on the psychological side of Benacerraf's chasm, while (3) and (4) constrain the metaphysical side.

What (1) rules out is that the subject's understanding of a normative predicate is so partial or internally incoherent that it cannot ground stable verdicts about the extension of her predicate 'N', even after ideal reflection and full descriptive information about her circumstances. This is an empirical claim about the subject's understanding of normative terms, which could turn out to be false. In effect, (1) requires that there be a stable upshot of reflective equilibrium that yields determinate verdicts about sufficiently many cases. This is a relatively uncontroversial, or at least widely accepted claim in contemporary metaethics: in addition to theorists like Peacocke and Jackson, Gibbard-style expressivists, hybridists, contextualists, relativists, and constructivists, for instance, regularly take the subject's ideal verdicts about the normative status of possible actions to demarcate an extension in all possible worlds (in some cases relative to some extra parameter marking a point of view or a context of use).

To appreciate this point, it's worth emphasizing what (1) does not say. First, (1) does not require that ideal reflective equilibrium suffice for deciding whether N or not-N is instantiated in every possible case. It may be that one's ideal criterion for identifying N fails to adjudicate some cases. If so, an intension assigned on the basis of your ideal verdicts will be undefined for those cases, and the instantiation of any property picked out will be indeterminate in those cases. It's plausible that most ordinary language predicates will have some degree of vagueness in their application conditions, so satisfying (1) should not require the elimination of all indeterminacy. Second, the constraint does not place any conditions on which factors – psychological, environmental, or social – explain why the upshot of empirically informed reflective equilibrium is sufficiently determinate. Some proponents of (IA) hold that conceptual understanding is constituted by a proper subset of "primitively compelling" or "obvious and central" beliefs or inferential dispositions, which guide one's ideal verdicts about the reference (Peacocke, Jackson). However, (IA) is not committed to any such psychologically real, causally efficacious states that guide one's use of a concept come what may. Contingent facts about human psychology or about one's actual linguistic community may be crucial to explaining why one's ideal, fully informed verdicts yield sufficiently determinate instantiation conditions for N. Determinacy need not be grounded in anything like an internal mental template for identifying the reference of 'N': determinacy may be grounded in contingent empirical facts about one's actual circumstances.^{xv}

Of course, (1) could be challenged. For instance, one might hold that the subject's ideal reflection will fail to yield any stable normative verdicts for any but the most trivial cases (e.g. that torturing babies for fun is morally wrong). If this is right, then (IA) will not suffice to vindicate a sufficiently determinate form of normative realism (on this point, see (Schroeter and Schroeter 2013, 7-8) which argues that normative terms tolerate less indeterminacy than many natural kind terms). In the final section of the paper, we'll return to this issue when considering a different metasemantic approach that supplements the upshot of ideal reflection with further facts to fix a determinate reference.

The supervenience claim in (2) is likewise confined to the epistemic perspective of the subject. From the perspective of the subject reaching ideal reflective equilibrium, the applicability of normative predicate 'N' cannot float free from the more basic descriptive facts: if two cases are descriptively indiscernible, then they must be indiscernible with respect to the applicability of 'N'. What (2) rules out is the idea that the subject would, on ideal reflection, be disposed to assume that normative properties are radically dualistic in the sense that they can vary independently of descriptive properties. Of course (2) is open to empirical challenge. But proponents of (ANR) will likely find the claim congenial. So we take the claim to be largely uncontroversial in the present context.

Condition (3) proposes a metaphysically lightweight conception of properties. David Lewis famously suggested that we need two sorts of properties in our metaphysics: sparse properties and plenitudinous properties (1984). Sparse properties are restricted to the most fundamental objective dimensions of similarity that can figure in basic causal-explanatory laws of nature and metaphysical theorizing, whereas plenitudinous properties are posited to capture all of the different features that we can talk about in natural language. For metasemantic purposes, the second notion is the most relevant one. After all, we're able to think and talk about much more than just the most fundamental features of reality like *being a photon* or *having negative charge*. So we need a metasemantics that can attribute the right extension to our talk about *being groovy* or *being grue*. The most liberal approach to properties is to simply identify a property with the set-theoretic construct that captures its instantiation conditions in every possible world. So the property of *rightness*, for instance, would be an infinite set that specified which action type *a* is right for every possible choice situation, as determined by a possible world *w*, a subject *s*, and a time *t*: [*<w1, s1, t1, a1>*, *<w1, s1, t2, a2>*, *<w1, s1, t3, a3>*, ..., *<w2, s2, t1, a4>*, *<w2, s2, t2, a1>*, ...]. The nice thing about the set-theoretic conception of properties is that it provides the resources for specifying any possible way of classifying things.

Condition (4) imposes a correspondingly thin condition on property identity: all it takes to be the same property is to have the same instances in all possible worlds. This is the most common way of individuating properties in the metaphysics literature. A key advantage of this approach

is that it allows for true property identities, like *being water = being H₂O* or *being a triangle = being a trilateral*. Finer-grained conceptions of property identity impose further constraints, such as playing the same role in causal or metaphysical explanations, or having the same cognitive significance in thought. Since explanation and cognitive significance can be accounted for on the psychological side of the metasemantic chasm we're interested in here, we will set them aside here.

With these conditions in place, the metasemantic constraint (IA) ensures that your use of a normative term like 'is right' picks out a unique property. If (1) and (2) are satisfied, then ideal reflective equilibrium will yield an intension (i.e. an extension in all possible worlds) which can be characterized enumeratively in purely descriptive terms. According to (3), this set-theoretic construct picks out a property. And (4) ensures that there is *only one* property picked out. Since (IA) tells us that the upshot of ideal reflective equilibrium is correct, it follows that there is a unique property picked out by your normative term 'is right'.

The important thing to notice is that (IA) provides a genuine bridge between psychology and metaphysics: between the subject's understanding of normative concepts and the properties those concepts pick out (if any). In order to explain reference determination, we need to start with a non-question-begging characterization of the relevant psychological states (e.g. the subject's first-order inferential and recognitional dispositions, her second-order dispositions to refine those dispositions, and her social or physical environment) and show how these facts secure a determinate property as reference. According to (IA), the crucial psychological fact is whether the subject is disposed to reach reflective equilibrium about how to classify cases if she were fully informed about her empirical circumstances (characterized in more basic descriptive vocabulary). (IA) simply defers to the subject's own ideal empirically informed judgments in determining the boundaries of the property her normative thoughts pick out. Notice, moreover, that this explanation of reference determination does not rely on any assumptions about the causal efficacy or mind-dependence of the properties in question. What's doing the work is the subject's psychological states and some very minimal metaphysical assumptions about properties.

In this section, we're focused on explaining how it's possible for reference to be fixed if we assume (ANR). Our goal here is not to defend (IA) against all possible objections, but to show how it might figure in a plausible response to (GIC) for (ANR). As we noted earlier, we take condition (2) to be largely uncontroversial in the present context, and we'll return to (1) below in section 7. In the remainder of this section, we'll focus on conditions (3) and (4), which concern the metaphysical side of Benacerraf's chasm.

Condition (4) has attracted a lot of attention in recent debates about normative naturalism. It seems to ground a quick argument for naturalism simply on the basis of global supervenience (Jackson 1998, 122-3). So some normative naturalists have sought to defend it (Jackson 1998, 125-8; Streumer 2008, 543-5; 2013 section 4), while many non-naturalists have sought to reject it by invoking supposed counter-examples (Shafer-Landau 2003, 93-5; FitzPatrick 2008, 199-200; Enoch 2011, 137-40; Parfit 2011, 296-7). So one might object that our metasemantic proposal rests on a question-begging assumption about the nature of properties that is biased towards naturalism.

However, this objection misconstrues the dialectic of our argument. Our motivation for proposing (4) is not that it can vindicate naturalism, but rather that it can play a crucial role in explaining univocal reference, and ultimately help resolve (GIC) for Autonomous Normative Realism. If properties are individuated more finely than necessary co-instantiation, then we need some account of how it's possible to univocally pick out just one among many different necessarily co-instantiated properties. First, at the metaphysical level, we need to know what objective metaphysical characteristics distinguish these different properties from each other. But more importantly for present purposes, we then need an explanation of how the psychological patterns of understanding and use of a predicate 'N' ensure that just one of these necessarily co-instantiated properties counts as the reference. Given that our ideal classificatory dispositions don't distinguish among them, (IA) cannot do this job. Given the different referential candidates are causally inert, no causal condition on reference can help. It's not obvious what else can do the job of securing univocal reference.

A different line of objection to our metaphysical assumptions is that the set-theoretic conception of properties posited by (3) is too thin to capture genuine properties. The worry is that an extension in all possible worlds is just a long list of action/circumstance pairs – so it doesn't capture the *similarity* among those instantiation conditions that's distinctive of a property (Scanlon 2014, 45). But this objection is not warranted when the instantiation conditions are generated by (IA). (IA) assigns an intension that captures the subject's ideally reflective and informed verdicts about which actions deserve to be classified together as right. So the instantiation conditions generated by (IA) will reflect an intuitive dimension of similarity, rather than a mere collection of instances. From the subject's point of view, the intension is not just a random collection of cases – it demarcates a subjectively discernible pattern of empirical similarity relevant to the target domain. Consider some other examples. From the point of view of the universe, the instances (IA) generates for expressions like 'is fashionable' or 'is a university' or 'is grue' may seem like arbitrary heterogeneous collections. But from the perspective of the user of these expressions, these collections reflect subjectively important commonalities in virtue of which different possible instances get included in the set. (IA) thus guarantees that the set-theoretic constructs assigned to our representations will share some subjectively discernible similarity relation. This is what we call a metaphysically *thin property*: an empirical similarity relation that is individuated by necessary co-instantiation.

Moreover, condition (3) – like condition (4) – is motivated by the search for a plausible answer to (GIC). If we limit ourselves to sparse properties, or endorse some other metaphysically restrictive conception of properties, we run the risk of reference failure for most of our predicates. For instance, if we were to require that properties must play a non-trivial causal role or that they capture a fundamental dimension of reality, then predicates like ‘is a university’ and ‘is grue’ arguably wouldn’t pick out properties. But clearly such terms can and do pick out thin properties. If we allow for thin properties, there is no risk that the world might fail to provide a property whose instantiation conditions correspond to the application conditions generated by ideal, empirically informed reflection. Condition (3) thus plays a dialectical role parallel to Balaguer’s appeal to plenitudinous entities in the mathematical case: it guarantees that there will be a thin property picked out by any predicate we would ideally take to have stable empirical application conditions.^{xvi}

In sum, our goal in this section has been to show how it’s possible to secure determinate reference to properties of the sort posited by the (ANR). To this end, we have proposed an influential metasemantic constraint (IA) together with a metaphysically lightweight conception of thin properties. We have also assumed that ideal, empirically informed reflection will support stable verdicts about the empirical applicability conditions of one’s normative terms. Together, we suggest, these claims provide one plausible route for resolving the metasemantic aspect of (GIC). All of these claims might be challenged, of course. However, the point we want to emphasize is that the proponent of (ANR) cannot simply reject these claims piecemeal without providing some alternative account of how determinate reference to the sort of properties she posits is achieved. If the normative realist finds any of these claims implausible, the challenge is to replace this answer to (GIC) with a better one.

4. Reliability

Let’s assume that our arguments have been successful and that (IA) can secure determinate reference to mind-independent, causally inert properties. Can (IA) also dispel worries about the reliability of the dispositions generating normative beliefs?^{xvii}

To clarify the types of reliability at stake we’ll first consider Field’s Nepalese village analogy. Suppose a subject has a large stock of beliefs about events in a causally isolated Nepalese village – such as the names and ages of its inhabitants, where they live, their political and romantic intrigues, and so on. Let’s call beliefs about the instantiation of such N-properties the subject’s ‘N-beliefs’.

To get a measure of the reliability of the subject's belief-forming dispositions, we need to consider the counterfactual correlation between changes in events in the isolated Nepalese village and changes in the subject's beliefs about them.

Counterfactual Correlation: If the distribution of N-properties were different then the subject's N-beliefs would be correspondingly different (*ceteris paribus*) to yield N-beliefs that are mostly true – or true relevantly more often than not.

Counterfactual Correlation tests whether the subject's first-order dispositions for identifying the instantiation of N-properties – her *N-criteria* – are sensitive to the actual instantiation of those properties. If the demographic, political, or romantic facts about the Nepalese villagers were to change, the subject's beliefs about them would change accordingly to maintain the frequency of true beliefs about these matters. In the Nepalese case, Field's point is that there seems to be no mechanism that could explain Counterfactual Correlation. In that case, there would be no satisfactory answer to (GIC): it's utterly mysterious what keeps our beliefs in line with changes in the distribution of properties in the domain.^{xviii}

Notice that Counterfactual Correlation says nothing about our ability to rationally improve our reliability about the target domain. As reflective epistemic agents, we critically assess and fine-tune our current N-criteria to increase the proportion true N-beliefs. So a further question one can ask about a subject's reliability is whether her reflective dispositions to revise her current N-criteria are in fact likely to increase their reliability.

Improvement: The subject's higher-order dispositions to refine her current N-criteria in the light of theoretical reflection and empirical inquiry would lead towards *more reliable* N-criteria (i.e. application dispositions that would yield a greater proportion of true beliefs, *ceteris paribus*).

Improvement gauges the subject's reflective abilities to get closer to the truth about N-properties. One notable feature of Field's thought experiment is that the subject is in no position to assess the reliability of her current N-criteria – much less to improve their reliability. So even if her dispositions to form beliefs about the Nepalese village were reliable, any reflective revision of those dispositions would be a shot in the dark – it would just as likely decrease her reliability as increase it. Improvement is particularly important in the normative and mathematical domains, since our epistemology in these cases largely turns on critical reflection and theorizing. But even in ordinary empirical domains like ascertaining the location of houses in one's home town, we ordinarily have sufficient reflective understanding of our current N-criteria to put us in a position to improve them through critical reflection – e.g. by improving our perceptual access or identifying more reliable witnesses. So an adequate answer to (GIC) will normally need to explain Improvement (or explain it away).

Now let's consider how Counterfactual Correlation and Improvement might be vindicated in the case of (ANR). Our suggestion is that the metasemantic constraint (IA) helps explain both

Improvement and the reliability of the dispositions generating our normative beliefs. It thus provides a plausible route to answering (GIC).

Let's start with Improvement. According to (IA), referential assignments must respect the upshot of ideal, empirically informed reflection about possible cases. This means that your reflective dispositions to refine your criteria for applying normative terms must be accurate – at least in the long run. In the short run, of course, there may be misleading evidence or lines of reasoning that lead you astray. But (IA) guarantees that the stable upshot of theoretical reflection and empirically informed inquiry will lead you to the truth about what it takes for a given object or property to be instantiated. This is a significant result for our attempt to address the General Integration Challenge for (ANR): (IA) provides *a secure deliberative route* to the truth about the instantiation of normative properties. This deliberative route is backed by a necessary metasemantic relation between the subject's current total understanding of normative concepts and the reference of those concepts. If the subject's understanding were sufficiently different that it would lead to a different pattern of ideal verdicts about possible cases, then the subject would eo ipso have a different reference and a different normative concept.

Next consider Counterfactual Correlation. In the normative domain, Counterfactual Correlation requires that one's criteria for coming to a verdict about whether a normative property is instantiated be sensitive to variations in the instantiation of those properties. If moral wrongness, for instance, had been differently distributed, our beliefs about its instantiation would have varied accordingly, so as to maintain mostly true beliefs about the existence of wrong actions in the world. Had Saddam Hussein promoted democracy among his people instead of gassing them and had George W. Bush engaged in diplomacy instead of rushing to war, for instance, we would not have judged their actions morally wrong.

By itself, (IA) does not ensure that your current criteria for coming to verdicts about which actions are right or wrong are reliable. It's consistent with (IA), for instance, that your current criteria for identifying witches leads you to systematically misclassify people as witches. You may currently be disposed to classify eccentric obstreperous old women as witches, even if full empirical information and ideal reflection would lead you to conclude (i) that there are no witches or (ii) that the only real witches are members of the Inquisition. So your actual N-criteria could be highly unreliable guide to the instantiation conditions of N-properties, and your N-beliefs could be mostly false.

Still, according to (IA) error theories or radical revisionist theories of N-properties will be true only if they accord with the upshot of reflective equilibrium. The relevant verdicts about the non-instantiation of N are true only if they would strike you as justified on the basis of your

current understanding of 'N' together with full base-level empirical information about your environment and ideal reflection. So if your current N-criteria are unreliable, it must be because you have other, more theoretical commitments about the reference that would strike you as more important to determining the nature of the reference at the point of ideal, empirically informed reflective equilibrium. (IA) thus provides a *conditional* guarantee of counterfactual correlation:

Conditional Counterfactual Correlation (CCC): If the subject's theoretical commitments about the reference don't require her to accept an error-theory or to radically revise her current N-criteria at the ideal limit of empirically-informed reflection, then her N-criteria are reliable (in the sense of Counterfactual Correlation).

One might worry that (CCC) is too weak to provide a plausible solution to (GIC). But in fact, merely conditional reliability is a feature of (IA), not a bug. In its unqualified form, Counterfactual Correlation conflicts with commonsense fallibilism about one's current N-criteria. As semantic externalists have emphasized through numerous examples, we could turn out to be ignorant or mistaken about the instantiation-conditions familiar objects, kinds, or properties. There might be good reasons to conclude that Gödel didn't discover the incompleteness theorem, that lemons aren't really yellow, that there are no human races, and so on (Burge 1979; Kripke 1980; Putnam 1970). Realists in particular should be open to these sorts of radical critique of one's current N-criteria. (IA) allows for the full range of fallibility that we, as rational epistemic agents, take ourselves to be subject to. By allowing radical revisions of one's current N-criteria in ideal reflective equilibrium, (CCC) reflects the subjects' ideal reflective commitments to their own fallibility.

Moreover, (CCC) is not a toothless constraint: it provides an important metasemantic link between one aspect of the subject's total understanding of 'N' (her current N-criteria) and the metaphysical instantiation conditions of N-properties. According to (IA), the subject's N-criteria can be systematically mistaken, but only if there are other theoretical aspects of her total understanding that she would find more important after ideal, empirically-informed reflection. For our purposes, the important point is that (IA) bridges Benacerraf's chasm between N-beliefs and N-properties: it dissolves the mystery of how our beliefs about mind-independent causally inert properties could be reliable. Our reliability about the instantiation of N-properties is explained at the metasemantic level, by tying reference determination to the upshot of ideal empirically informed reflective equilibrium. (IA) thus directly ensures the truth of Improvement, and it ensures a suitably hedged versions of Counterfactual Correlation. In contrast to the Nepalese village case, there is no need to posit any mysterious brute correlation between our normative beliefs and the realm of mind-independent, causally inert properties.

Some proponents of (ANR) might be tempted to endorse (CCC) while remaining neutral about (IA). This position might seem attractive because it avoids any commitment to a strong

metasemantic thesis positing infallibility at the ideal limit of empirically informed reflection. But it is crucial to see that what's doing the work in the proposed response to (GIC) is the metasemantic principle (IA), not (CCC) by itself. (CCC) merely asserts the conditional reliability of one's N-criteria without explaining why the conditional holds. In contrast, (IA) affords a coherent explanation: it explains how the contents of our N-beliefs depend on our patterns of understanding, and how this link ensures an epistemic route to the truth about the instantiation-conditions of N-properties. When (CCC) is derived from (IA), it constitutes part of a cogent response to the Generalized Integration Challenge.^{xix}

Let's review the argument. In the previous section, we argued that (IA) can secure determinate reference across Benacerraf's chasm between our psychological states and the causally inert, mind-independent properties they pick out. In this section, we invoked (IA) to explain the reliability of our normative beliefs. First, we explained how (IA) can vindicate the reliability of our higher-order dispositions to correct our first-order criteria for classifying actions as right. Second, we argued that (IA) explains the conditional reliability of these first-order criteria. This explanation of conditional reliability does not appeal to any brute correlation between our normative beliefs and normative properties; instead it rests on a substantive explanation of how Benacerraf's chasm can be bridged by (IA).

5. Causal inertia and mind-independence

We still need to address some important unfinished business. In section 3, we claimed that (IA) is consistent with the causal inertia and mind-independence of normative properties. In this section, we'll consider challenges to these claims. Can (IA) really pick out mind-independent, causally inert properties of the sort posited by Autonomous Normative Realism?

In section 1, we suggested that a property is causally inert iff it doesn't figure essentially in our best causal explanation of events. A wide range of empirical properties are causally inert in this sense: e.g. *being a chair*, *being fashionable in Paris in 1450*, *being grue*, etc. Their causal inertia does not, of course, prevent us from appealing to these properties in loose everyday causal explanations. Even if being a chair doesn't figure in our ultimate causal theory of the universe, we can explain why someone was knocked down noting that he was hit with a chair. These explanations are successful because the property of *being a chair* is realized in virtue of the instantiation of more fundamental causally relevant properties, like being a medium-sized rigid object within a certain mass range exhibiting certain broad topological features. So we can cite *being a chair* in commonsense 'program explanations' that constrain the range of fundamental properties of the object which would figure in our ultimate causal explanation of the event (Jackson and Pettit 1990). Similar points can be made in the normative case: even if the property of *being evil* isn't indispensable in our causal-explanatory theory of events, it still demarcates a pattern of empirical similarity that can be useful in everyday causal explanations.

It's controversial, of course, whether normative properties like *being right* or *being evil* really are causally inert in this sense. Some Cornell realists seem to hold that normative properties do play a crucial causal-explanatory role (Sturgeon 1985). But we need not enter into this dispute here. (IA) is consistent with the possibility that the property picked out figures essentially in our best causal explanations. But our point is simply that *if* proponents of (ANR) are right that normative properties are causally inert in the relevant sense, (IA) can still single them out as the reference of our normative terms.

One might worry, however, that some proponents of (ANR) take normative properties to be causally inert in a more robust sense. David Enoch (2011), for instance, takes the analogy with mathematical Platonism very seriously. Mathematical Platonism seems to posit a realm of abstract objects that are wholly outside the causal order, whose existence is entirely independent of the instantiation of any empirical properties. Perhaps some proponents of (ANR) hold that normative properties are similar: they're denizens of an abstract 'third realm' or a *sui generis* dimension of normative reality that's wholly outside the causal order. In that case, one might argue that our response to (GIC) radically underestimates the difficulty of bridging Benacerraf's chasm.

However, the analogy does not hold. Even non-naturalists like Enoch accepts that (i) ordinary empirical events actually instantiate normative properties. So normative properties are not denizens of a purely abstract 'third realm' wholly isolated from the empirical domain. Moreover, these non-naturalists must also accept that we do cite normative properties in loose everyday causal explanations, and that these explanations are often informative. So (ii) the instantiation of normative properties is at least roughly correlated with an empirical dimension of similarity that can figure in loose causal explanations. These two commitments suffice for the type of minimal role in loose everyday causal explanation posited above for normative properties. So even non-naturalists like Enoch do not take normative properties to be radically cut off from the realm of loose everyday causal explanation in the way that Platonists take mathematical objects and properties to be.

Our second question is whether (IA) is really consistent with the Autonomous Normative Realist's commitment to the mind-independence of normative properties. We'll consider two lines of objection to this idea.

The first objection is that (IA) is committed to constructivism about normative properties, and so it cannot vindicate (ANR):

(IA) is committed to metaphysical constructivism about normative properties, since it makes normative facts depend on the subject's ideal normative judgments.

Although this objection is often pressed in discussion, it can be dealt with quickly since it rests on a simple conflation. The problem with the objection is that it fails to distinguish between metasemantic and metaphysical mind-independence. Consider the demonstrative expression 'this stone'. The object this expression refers to is metaphysically mind-independent: its existence does not depend on the subjective attitudes or commitments anyone has towards it. But the fact that the expression 'this stone' refers to one particular object, *o*, among all the possible objects, kinds, or properties it might pick out, depends on complex facts about the way those words are understood and used by the speaker. In particular, the reference relation depends in part on the speaker's understanding of 'stone' and the speaker's intention to focus the audience's attention on a particular object that has been made salient by the conversational context. Thus the truth conditions for an utterance involving the expression 'this stone' depend in part on the attitudes and commitments of the speaker towards *o*.^{xx}

The type of mind-dependence involved in (IA) is *metasemantic*: if your predicate 'is right' were associated with a sufficiently different pattern of understanding, its reference would be different than it in fact is. Thus, which property (if any) is picked out by the concept expressed by your predicate 'is right' depends at least in part on your associated mental states.

In contrast, metaphysical mind-dependence concerns the instantiation conditions of the property you actually pick out, independently of how reference to that property is secured. A metaphysical constructivist might claim, for instance, that what it is for an action to be right *just is* for there to be someone who has mental state *M* towards that action (e.g. for there to be a community of type *T* that's disposed to reflectively endorse that action). Such metaphysical claims entail necessary *counterfactual co-variation* of the properties: there is no possible world in which rightness is instantiated and attitude *M* is not.

These two types of dependence, metasemantic and metaphysical, should be sharply distinguished. The first posits a constraint on the semantic contents assigned to mental states, whereas the second posits a necessity relation between the instantiation of a given object, kind or property and the instantiation of certain mental states. (IA) involves no metaphysical commitment to the mind-dependence of the objects, kinds or properties picked out. (IA) is consistent with reference to wholly mind-independent features of reality, such as water or stones. Moreover, the type of metasemantic mind-dependence posited by (IA) is an inevitable consequence of Minimal Charity: if mental contents are to play their role in normative assessment and prediction and explanation, semantic assignments must depend in part on subjects' patterns of understanding.^{xxi}

The second objection is that (IA) doesn't yield the kind of mind-independence needed to vindicate the distinctive *objective purport* of normative discourse. According to (IA), two individuals or communities co-refer with their use of a normative predicate such as 'is right' only if they would ideally converge in all their ideal verdicts about which actions are right in every possible worlds. If there is any divergence in their ideal verdicts, however slight, it follows that they use 'is right' to pick out distinct properties, so they cannot use that term to anchor strict logical agreement or disagreement with each other.

According to robust normative realists like David Enoch, this is not an acceptable result. (IA) makes normative truth hostage to psychological fortune and opens the door to relativism. Given that human nature allows for considerable variation in their practical and theoretical dispositions, it is not difficult to imagine communities (actual or hypothetical) with whom we would fail to ideally converge on normative questions. But normative truths are "perfectly objective, universal, absolute" (Enoch 2011, 1). So (IA) is unable to deliver the type of universality proponents of Autonomous Normative Realism such as Enoch take to be central to normative discourse. To vindicate universality of normative discourse, we'd need a response to (GIC) that makes the normative domain more "robustly" mind-independent.^{xxii}

To see what's at stake here, we need look more closely at Enoch's contention that normative truths are "perfectly objective, universal, absolute". The claim is not simply that all competent English speakers pick out the same property with their term normative predicates like 'is right'. This form of context-invariant cognitivism leaves open the possibility that different linguistic communities may use their normative terms to pick out different properties – a situation Enoch's wants to rule out. Suppose, then, that different communities have normative concepts that play a role in action and practical deliberation that's similar to the role played by the concept we express with 'all told right' in English. Enoch's suggestion seems to be this:

Semantic Uniqueness (SU): Any linguistic community that has an all told normative concept (i.e. a concept that plays a similar role in deliberation and guidance of action as the concept we express with 'all told right') must pick out the very same property, which determines the truth-conditions for claims involving that concept.

Semantic Uniqueness is a generalization of context-invariant cognitivism. Whereas context-invariant cognitivism is restricted to a particular concept, such as the concept expressed by 'all told right' in English, (SU) targets any concept that plays a buck-stopping role in deliberation and action.^{xxiii}

In response to Enoch, the first point to notice is that (SU) has nothing to do with the mind-independence of normative properties. (SU) would be satisfied even if all buck-stopping normative concepts secured reference to the very same mind-dependent property, *being*

beloved by the gods. (SU) makes a semantic claim about *sameness* of reference, not a metaphysical claim about the nature of the property picked out.

The main problem with (SU), however, is a familiar one: it's the problem of radical and persistent normative disagreement. As several authors have noted, the main challenge radical and persistent disagreement poses to metaethical theories traditionally labeled as 'realist' is a metasemantic one (Tersman 2006, 83-106; Björnsson 2012; Schroeter and Schroeter 2013). The challenge is to provide a metasemantic theory that can vindicate co-reference, despite radical divergence in individuals' ideally reflective verdicts about the applicability conditions of normative predicates like 'is right'. There are ways of explaining such co-reference within a given linguistic community, by appealing to individual speakers' intentions to coordinate with a communal linguistic practice (van Roojen 2010; Schroeter and Schroeter 2009; 2014). But (SU) posits co-reference between different linguistic communities: all possible subjects who have a concept that plays a buck-stopping role in deliberation and action must co-refer. It's not obvious what could secure co-reference in this case.

We believe that no plausible metasemantic theory can vindicate co-reference across the radically divergent patterns of understanding and use that might be associated with a concept that plays the buck-stopping role (Schroeter and Schroeter 2013, in press a). Our basic worry is simple: semantically guaranteed co-reference cannot be secured unless it can be explained on the basis of the conditions for conceptual competence. But the buck-stopping role does not include constraints that suffice to secure reference to an empirically instantiated property. Without some empirical commitments as input, the metasemantic theory cannot yield empirical applicability-conditions as output. We won't repeat our arguments here.

The important point for present purposes is that anyone who accepts (SU) owes us an explanation, if they are to meet (GIC): they must provide a plausible metasemantics that can vindicate co-reference simply on the basis of the buck-stopping role of all told normative concepts. We doubt this can be done. But if it can, the further question is then whether their favored metasemantics provides proponents of (ANR) with a response to (GIC) that is superior to the one we have sketched in this paper.

6. Taking stock

We've introduced the Generalized Integration Challenge as a criterion of explanatory adequacy for a philosophical account of any domain. We then used (GIC) to help address epistemic puzzles faced by Autonomous Normative Realism. Taking metasemantics into account can help establish an internal relation between the psychological realm of normative judgments and the metaphysical realm these judgments are about. Providing a metasemantic theory that explains

how our psychological states manage to pick out specific properties constitutes a first step towards explaining how we can have knowledge about mind-independent causally inert normative properties. To fulfill this explanatory promise, we appealed to an influential, independently motivated metasemantic constraint on reference determination, Ideal Accessibility. Ideal Accessibility ensures that we have an epistemic route to the applicability conditions of our referential concepts, and it provides a straightforward explanation of how normative concepts can have as their reference properties that are mind-independent and causally inert. On this proposal, the normative property of *all told rightness* posited by Autonomous Normative Realism is an empirical dimension of similarity that the subject would deem relevant after ideal, empirically informed reflective equilibrium.

Some metaethicists have advocated normative realism on the basis of metasemantic theories of reference determination. Christopher Peacocke's *moral rationalism*, for instance, is grounded in his foundationalist account of concept possession. Moral concepts, Peacocke argues, involve an implicit conception of certain basic moral principles and all other moral truths can be derived from these basic principles on the basis of a priori reasoning (Peacocke 2004, Ch. 7). Frank Jackson and Philip Pettit's *moral functionalism* is also motivated by metasemantic considerations. Their account of reference determination is inspired by holistic 'contextualist definitions' of theoretical terms: the reference is determined as whatever properties best satisfy a subject's implicit 'folk theory' of the moral domain (Jackson 1998; Jackson and Pettit 1990; Lewis 1970). Our proposal is consistent with either approach, but it avoids some of the specific commitments to analytic truths about the empirical applicability conditions for normative concepts that are characteristic of these approaches. The key difference is that (IA) does not place any strong commitments on competence conditions for concepts. In particular, it's consistent with (IA) that competence with the same concept is determined in part by causal-historical relations to a historically-extended representational tradition. As we've argued elsewhere, this means that mere conceptual competence may not put one in a position to recognize correct verdicts about the applicability conditions of one's concepts on the basis of individualistic a priori reflection (Schroeter and Schroeter 2014).

Although our proposal to solve (GIC) for (ANR) is motivated by metasemantic considerations, it has important metaphysical consequences that distinguish it from the two main strands of Autonomous Normative Realism currently on the market. (IA) would generate ordinary empirical dimensions of similarity as the reference of a normative concept. We call the resulting metaphysical position *Deflationary Normative Naturalism* (DNN). (DNN) qualifies as a version of (ANR), since involves no commitment to normative properties being causally important or constituted by mental states. But because (DNN) is generated by Ideal Accessibility, the conceptually competent subject must have an epistemic route to discovering the instantiation conditions of normative properties through ideal empirically-informed reflection. In this sense, normative properties are the 'ontological shadows' of reflective equilibrium. This metasemantic commitment distinguishes (DNN) from Robust Realism, which holds that the instantiation conditions of normative properties are determined by independent metaphysical facts that may

outrun competent subjects' ideal reflective verdicts. (DNN) is also distinct from the anti-metaphysical brand of normative non-naturalism: whereas (DNN) simply identifies normative properties with ordinary empirical dimensions of similarity, anti-metaphysical proponents of (ANR) like Scanlon and Parfit reject naturalism and posit a metaphysical distinction between normative properties and necessarily co-instantiated empirical properties. So anti-metaphysical proponents of (ANR) also have more inflationary metaphysical commitments than (DNN).

Our goal in this paper, however, was not to argue that proponents of (ANR) should favor naturalism because it involves lesser metaphysical commitments. Naturalism is a by-product of our solution to (GIC). As we emphasized earlier, the debate between naturalism and non-naturalism in the normative domain should turn, not on general metaphysical preferences, but on which approach provides the best solution to (GIC).

We'd also like to stress that our proposed solution to (GIC) in the normative domain will not carry over straightforwardly to Platonism in the mathematical domain. Our defense of (DNN) hinges on the fact that normative properties are instantiated by ordinary empirical events. In contrast, the Platonist takes mathematical properties to be instantiated not by empirical entities, but by a realm of purely abstract entities. So the Platonist cannot appeal to (IA) to vindicate reference to mathematical objects and properties.

7. Extending the metasemantic strategy?

We have sought to show that *if* (IA) is a valid metasemantic constraint, then metasemantics can provide the key to resolving (GIC) for (ANR). However, not every theorist will accept (IA). For instance, proponents of (ANR) might question the psychological Determinacy assumption. One might doubt that an individual's ideal, empirically-informed reflection will ground determinate verdicts applying normative concepts in any but the most trivial cases. Or one might grant determinate verdicts, but doubt that they will be stable over time or between individuals. If such worries are borne out, (IA) won't fix a stable reference for seemingly stable normative concepts.

Recently, a number of metaethicists have proposed *reference magnetism* (RM) as a potential solution to this sort of worry about determinacy and reference stability (van Roojen 2006; Dunaway and McPherson 2016; Williams Ms.). The idea is to follow David Lewis in supplementing an internal charity constraint with a purely external constraint on reference determination (Lewis 1983; 1984). Lewis's suggestion, as standardly understood, is that certain features of the world are objectively better referential candidates than others. The correct interpretation, then, is one which best balances a *charity constraint* (e.g. maximizing true beliefs within the subject's total ideal theory) against an *eligibility constraint* (e.g. maximizing relative eliteness among the total referential assignments). On Lewis's account, perfect eliteness is explained in terms of the fundamental joints of nature, and relative eliteness is determined by how easy it is to construct a reductive definition for a target domain

in terms of perfectly natural properties. Moreover, Lewis's account is *holistic*: the interpretation of each representational state in an individual's repertoire depends on the totality of her cognitive dispositions. So the semantic content of every word or concept must be fixed simultaneously. It's worth emphasizing that this holistic assumption was not in place in our discussion of (IA). So far, we have assumed that the semantic content of non-normative terms used to describe possible cases is fixed independently of normative terms, and the reference of normative terms can then be fixed by tying them to descriptions of possible cases.

By introducing an external element and a holistic account of reference determination, Lewisian reference magnetism allows the proponent of (ANR) more flexibility about the type of charity constraint involved in securing reference. An eligibility constraint could secure referential assignments that make at least some of a subject's ideally reflective, empirically-informed normative judgments come out false – a position many normative realists hope to vindicate.^{xxiv} It's worth asking whether (RM) might help solve (GIC) for autonomous normative realism.

The short answer is that it all depends on the details. The bulk of this paper has been devoted to articulating this detail for one influential metasemantic proposal, (IA). A similarly detailed articulation of (RM) would be needed to assess whether the approach can provide a metasemantic solution to the generalized integration challenge for (ANR). Obviously, we can't hope to undertake this task here. We'll simply make a few programmatic remarks about the prospects and challenges for a reference magnetic solution to (GIC). Given that (RM) invokes a charity constraint as a crucial component of the metasemantics, our discussion of (IA) can provide a useful reference point for assessing how such an account might address (GIC).

If we want (RM) to vindicate something like Conditional Counterfactual Correlation and Improvement, we will need a charity constraint that goes significantly beyond Minimal Charity. One possibility would be for (RM) to simply adopt (IA) as the charity constraint while maintaining that it must be balanced against an eligibility constraint.^{xxv} If (RM) is matched with a thin account of properties, the approach can explain (CCC) and (Improvement) in a similar manner to (IA). The difference is that (IA) ensures perfect accuracy of ideal empirically informed reflective equilibrium, whereas reference magnetism allows for mistakes at this ideal limit. As a consequence (IA) secures greater reliability along both dimensions, Improvement and Counterfactual Correlation. The more weight accorded to the eligibility constraint over the charity constraint, the less power the magnetic metasemantics has to explain the reliability of our normative beliefs.^{xxvi}

But it's important to note that giving too much weight to the eligibility constraint is implausible on independent grounds. Suppose that we stipulatively introduce a new normative term, 'Benthamite right' and 'Kantian right', to capture two different conceptions of moral rightness. The patterns of understanding associated with these terms overlap considerably and so the ideal verdicts about cases will also largely overlap. If (RM) imposes an eligibility constraint that secures co-reference despite

rationally irreconcilable disagreements between committed Kantians and Benthamites about the property picked out by ‘morally right’, then it seems it must also secure co-reference between our two theoretical terms that are stipulatively defined to capture these competing views. Despite our best intentions, then, we cannot use these terms to refer to distinct properties with these terms. Thus, an overly strong reference magnetism may make it impossible to talk about distinct, but closely related properties (Sundell 2012; Schroeter and Schroeter 2013; Williams Ms.). Such a metasemantic theory simply assigns the wrong reference.

The moral, we suggest, is that a plausible version of (RM) cannot invoke an eligibility constraint that would significantly override subjects’ ideal verdicts about cases. Reference magnetism will end up respecting a charity constraint much closer to (IA) than to (Minimal Charity). This is welcome news for solving (GIC), since it suggests that the proponent of (RM) can borrow the general metasemantic strategy we illustrated with (IA): a strong charity constraint explains the reliability of normative beliefs even when normative properties are autonomous and causally inert.

Of course, it’s an open question whether a plausible version of (RM) can be spelled out. But that’s not our concern here. Our aim is simply to highlight how metasemantics could potentially help address (GIC) for Autonomous Normative Realism.^{xxvii}

References

- Balaguer, Mark. 1998. *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.
- Benacerraf, Paul. 1973. Mathematical Truth. *Journal of Philosophy* 19: 661–79.
- Bengson, John. 2015. Grasping the Third Realm. *Oxford Studies in Epistemology* 5.
- Björnsson, Gunnar. 2012. Do 'Objectivist' Features of Moral Discourse and Thinking Support Moral Objectivism? *Journal of Ethics* 16: 367–93.
- Burge, Tyler. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* 4: 73–121.
- . 1993. Concepts, Definitions, and Meaning. *Metaphilosophy* 24: 649–65.
- Carnap, Rudolf. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.
- Chalmers, David. 2004. Epistemic Two-Dimensional Semantics. *Philosophical Studies* 118: 153–226.
- . 2006. The Foundations of Two-Dimensional Semantics. In *Two-Dimensional Semantics: Foundations and Applications*, ed. M. Garcia-Carpintero and J. Macia, 55–140. Oxford: Oxford University Press.
- . 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chalmers, David and Frank Jackson. 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review* 110: 315–61.
- Chrisman, Matthew. 2015. *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*. Oxford: Oxford University Press.
- Chudnoff, Elijah. 2013. *Intuition*. Oxford: Oxford University Press.
- Clarke-Doane, Justin. 2012. Morality and Mathematics: The Evolutionary Challenge. *Ethics* 122: 313–40.
- . 2014. Moral Epistemology: The Mathematics Analogy. *Noûs* 48: 238–55.
- . 2015. What is the Benacerraf Problem? In *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*, ed. F. Pataut. Dordrecht: Springer.
- Davidson, Donald. 1973. Radical Interpretation. *Dialectica* 27: 313–28.
- Dennett, Daniel. 1981. Intentional Systems. In *Brainstorms*, 3–22. Cambridge: Bradford Books.
- Devitt, Michael. 1991. *Realism and Truth*. 2nd ed. Princeton: Princeton University Press.
- Dowell, J.L. 2015. The Metaethical Insignificance of Moral Twin Earth. *Oxford Studies in Metaethics* 11.

- Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge MA: MIT Press.
- . 1988. *Explaining Behavior*. Cambridge, MA: Bradford, MIT Press.
- Dummett, Michael. 1978. *Truth and Other Enigmas*. London: Duckworth.
- . 1991. *The Logical Basis of Metaphysics*. Cambridge MA: Harvard University Press.
- Dunaway, Billy and Tristram McPherson. 2016. Reference Magnetism as a Solution to the Moral Twin Earth Problem. *Ergo* 3: 639–79.
- Enoch, David. 2010. The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It. *Philosophical Studies* 148: 413–38.
- . 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Field, Hartry. 1989. *Realism, Mathematics and Modality*. Oxford: Basil Blackwell.
- FitzPatrick, William. 2008. Robust Ethical Realism, Non-naturalism and Normativity. *Oxford Studies in Metaethics* 3.
- . 2014. Why There is No Darwinian Dilemma for Ethical Realism. In *Challenges to Moral and Religious Belief: Disagreement and Evolution*, ed. M. Bergman and P. Kain, 237–55. Oxford: Oxford University Press.
- Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
- Frege, Gottlob. 1892. On Sinn and Bedeutung. In *The Frege Reader*, ed. M. Beany, 151–71. Oxford: Blackwell (1997).
- Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Goldman, Alvin and Bob Beddor. 2015. Reliabilist Epistemology. In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, Frank and Philip Pettit. 1990. Program Explanation: A General Perspective. *Analysis* 50: 107–17.
- . 1995. Moral Functionalism and Moral Motivation. *Philosophical Quarterly* 45: 20–40.
- Joyce, Richard. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kaplan, David. 1989. Demonstratives. In *Themes from Kaplan*, ed. J. Almog, J. Perry and H. Wettstein, 481–563. Oxford: Oxford University Press.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

- Lewis, David. 1970. How to Define Theoretical Terms. *Journal of Philosophy* 67: 427–46.
- . 1974. Radical Interpretation. *Synthese* 23: 331–44.
- . 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61: 343–77.
- . 1984. Putnam's Paradox. *Australasian Journal of Philosophy* 62: 221–36.
- Linnebo, Øystein. 2006. Epistemological Challenges to Mathematical Platonism. *Philosophical Studies* 129: 545–74.
- Millikan, Ruth Garrett. 1984. *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Neander, Karen. 2006. Content for Cognitive Science. In *Teleosemantics. New Philosophical Essays*, ed. G. Macdonald and D. Papineau, 167–94. Oxford: Clarendon Press.
- Papineau, David. 1987. *Reality and Representation*. New York: Blackwell.
- Parfit, Derek. 2011. *On What Matters*. Oxford: Oxford University Press.
- Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.
- . 1993. How Are A Priori Truths Possible? *European Journal of Philosophy* 1: 175–99.
- . 1999. *Being Known*. Oxford: Oxford University Press.
- . 2004. *The Realm of Reason*. Oxford: Oxford University Press.
- Pérez Carballo, Alejandro. 2015. Semantic Hermeneutics. In *Metasemantics: New Essays on the Foundations of Meaning*, ed. A. Burgess and B. Sherman, 119–46. Oxford: Oxford University Press.
- Pietroski, Paul. 1992. Intentionality and Teleological Error. *Pacific Philosophical Quarterly* 73: 267–82.
- Plunkett, David and Tim Sundell. 2013. Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint* 13: 1–37.
- Putnam, Hilary. 1970. Is Semantics Possible? In *Language, Belief and Metaphysics*, ed. H. E. Kiefer and M. K. Munitz, 50–63. Reprinted in (Putnam 1975): 139–152. New York: SUNY Press.
- . 1981. *Reason, Truth and History*. New York: Cambridge University Press.
- . 1983. *Realism and Reason: Philosophical Papers, vol 3*. Vol. 3, *Philosophical Papers*. Cambridge: Cambridge University Press.
- Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge MA: MIT Press.
- Ridge, Michael. 2014. *Impassioned Belief*. Oxford: Oxford University Press.

- Russell, Bertrand. 1905. On Denoting. *Mind* 14: 479–93.
- Scanlon, T.M. 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schechter, Joshua. 2010. The Reliability Challenge and the Epistemology of Logic. *Philosophical Perspectives* 24: 437–64.
- Schroeter, Laura. 2017. Two-Dimensional Semantics. In *Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta.
- Schroeter, Laura and François Schroeter. 2009. A Third Way in Metaethics. *Noûs* 43: 1–30.
- . 2013. Normative realism: co-reference without convergence? *Philosophers' Imprint* 13: 1–24.
- . 2014. Normative Concepts: A Connectedness Model. *Philosophers' Imprint* 14: 1–26.
- . in press a. Metasemantics and Metaethics. In *The Routledge Companion to Metaethics*, ed. T. McPherson and D. Plunkett. New York and London: Routledge.
- . in press b. Reasons and Justifiability. In *The Many Moral Rationalism*, ed. F. Schroeter and K. Jones. Oxford: Oxford University Press.
- Setiya, Kieran. 2012. *Knowing Right From Wrong*. Oxford: Oxford University Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Clarendon.
- Silk, Alex. 2015. How to be an Ethical Expressivist. *Philosophy and Phenomenological Research* 91: 47–81.
- Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics* 9: 315–32. Reprinted in Stalnaker (1999a): 78–95.
- . 1999. *Context and Content*. Oxford: Oxford University Press.
- Stampe, Denis. 1977. Towards a Causal Theory of Linguistic Representation. In *Midwest Studies in Philosophy*, ed. P. French, H. K. Wettstein and T. E. Uehling, 42–63. Minneapolis: University of Minnesota Press.
- Street, Sharon. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127: 109–66.
- Streumer, Bart. 2008. Are there Irreducibly Normative Properties? *Australasian Journal of Philosophy* 86: 537–61.
- . 2013. Why There Really Are No Irreducibly Normative Properties. In *Thinking About Reasons: Essays in Honour of Jonathan Dancy*, ed. D. Bakhurst, B. Hooker and M. Little. Oxford: Oxford University Press.

- Sturgeon, Nicholas. 1985. Moral Explanation. In *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, ed. D. Copp and D. Zimmerman. Totowa NJ: Rowman & Allanheld.
- Sundell, Timothy. 2012. Disagreement, Error, and an Alternative to Reference Magnetism. *Australasian Journal of Philosophy* 90: 743–59.
- Tersman, Folke. 2006. *Moral Disagreement*. Cambridge: Cambridge University Press.
- . 2015. Reliability and Necessary Truth.
- Turri, John. 2015. Unreliable Knowledge. *Philosophy and Phenomenological Research* 90: 529–45.
- van Roojen, Mark. 2006. Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument. *Oxford Studies in Metaethics* 1.
- . 2010. Moral Rationalism and Rational Amoralism. *Ethics* 120: 495–525.
- Väyrynen, Pekka. 2013. *The Lewd, the Rude and the Nasty*. Oxford: Oxford University Press.
- Williams, J.R.G. Ms. Normative Reference Magnets.
- Yalcin, Seth. 2014. Semantics and Metasemantics in the Context of Generative Grammar. In *Metasemantics: New Essays on the Foundations of Meaning*, ed. A. Burgess and B. Sherman, 17–54. Oxford: Oxford University Press.

ⁱ Just as (IC) does not rule out metaphysical anti-realism or epistemic skepticism about a given domain, (GIC) does not rule out the possibility that the domain of discourse has no determinate semantic content. If the correct metasemantic theory fails to assign any determinate semantic content for ‘N’, then a full philosophical account must provide a metaphysics and epistemology that fits with this theory: e.g. we need a metaphysical account of the phenomena that give rise to the discourse and an epistemological theory of the justificatory status of beliefs in that domain.

ⁱⁱ This broad sense of metasemantics is consistent with Peacocke’s usage in his discussions of the Integration Challenge and of apriori knowledge (Peacocke 1993; 1999). Yalcin would call an account of the determination of thought content for thought a *theory of intentionality*, which he takes to be orthogonal to his narrow construal of *metasemantics*, which explains how compositional principles are determined by linguistic competence (2014). The narrow understanding of metasemantics is of little direct relevance to an integrated philosophical account of particular domains such as mathematics or morality. What’s of real interest outside of linguistic theory is the relation between our psychological states and the substantive semantic contents they express.

ⁱⁱⁱ This consensus can be challenged: one might deny there is any stable content expressed by the use of normative expressions like ‘is right’. But this nihilism about the meaning of normative expressions must then explain the apparent discipline of our communicative and reflective practices. For a general overview of the role of metasemantics in metaethics, see (Schroeter and Schroeter in press a).

^{iv} In recent work, William FitzPatrick suggests that normative properties might have causal efficacy (FitzPatrick 2014), so we'll take David Enoch as our paradigm example of Robust Realism here.

^v Similarly, challenges to normative realism grounded in the *etiology* of normative beliefs (so-called "evolutionary" challenges to normative realism, (Gibbard 2003; Joyce 2006; Street 2006) would lose much of their force if there is an explanation of the reliability of our normative beliefs. So the challenge to explain how it is that we have reliable normative beliefs (assuming the truth of (ANR)) is more basic than the challenge to explain how it is that we came to have reliable normative beliefs (for relevant discussion, see Enoch 2010; Bengson 2015).

^{vi} On this point, see for instance (Linnebo 2006; Enoch 2010; Bengson 2015).

^{vii} We will use the term 'concept' to mark representational state types, rather than abstract entities. On this way of speaking, concepts are representational states that *have* semantic contents: e.g. they may represent objects or properties, or they may have non-representational inferential or expressive contents. This use coheres with the use of the term 'concept' in psychology and with that of many philosophers (Peacocke 1992; Burge 1993). However, nothing in this paper turns on the decision about how to use the term 'concept'. Our focus here is on the psychological states that suffice for competence with the semantic content of an expression.

^{viii} Conceptual understanding is intended to include any *psychologically real* state that might figure of the subject's current implicit understanding of the topic in question. But it does not include states that would require further reasoning to crystallize. Moreover, the states construed as accurate must involve *substantive commitments* about the topic in question. Minimal Charity will not be satisfied simply by vindicating the subject's acceptance of tautologies like 'what's right is right' or pragmatically self-verifying claims like 'Gödel is the man I call 'Gödel'".

^{ix} The first class of theorists includes traditional descriptivists, who hold that the subject implicitly grasps necessary and sufficient conditions for being the reference (Carnap 1947; Frege 1892; Russell 1905). It also includes neo-descriptivists and 2D theorists, who hold the weaker thesis that the subject's current understanding puts her in a position to identify necessary and sufficient conditions for being the reference on the basis of information about her environment (Peacocke 1992; Jackson 1998; Chalmers 2006). And it includes interpretivists, who hold that the correct semantic interpretation depends on construing the subject's cognitive economy as maximally rational given her empirical circumstances (Davidson 1973; Lewis 1974; Dennett 1981). So the correct interpretation of a given expression must be rationally justifiable on the basis of true beliefs and truth-conducive methods the subject associates with the term (perhaps in conjunction with empirical facts about her environment).

^x Other naturalistic theories focus on explanations that turn on the causal-historical origins of the subject's current attitudes and dispositions about a topic. Teleosemantic theories, for instance, ground semantic interpretation in natural selection for certain cognitively advantageous functions (Millikan 1984; Papineau 1987; Dretske 1988). Roughly a teleosemanticist holds that 'water' refers to H₂O iff the explanation of why tokens of that type were reproduced (within an individual or within a population) appeals to the fact that past tokens were caused by H₂O. Teleosemantics requires there to be veridical states in the historical period of natural selection: you count as referring to water only if you (or your ancestors) had true 'water' beliefs in the past. This means that Minimal Charity will be satisfied over the history of use of representations of that type. However, Ruth Millikan has emphasized that the historical aspect of her account allows for the possibility that an individual's current beliefs and dispositions about water may be systematically mistaken. This would violate Minimal Charity for the individual. Many theorists have argued that this is a reason to reject this version of

teleosemantics (Pietroski 1992; Neander 2006) – for it is not useful to explaining an individual's reasoning and action.

^{xi} A different 'direct realist' strategy for crossing the chasm has been recently proposed by John Bengson and Elijah Chudnoff: they suggest that normative properties like moral rightness are partly constitutive of normative judgments (Bengson 2015; Chudnoff 2013). This constitution relation is a much stronger relation than a metasemantic aboutness relation linking mental states with the normative properties they represent: the idea is that the mental state is somehow built out of the normative property. The direct realist position is reminiscent of the Aristotelian account of perceptual content, in which the forms instantiated by external objects (e.g. triangularity) are also instantiated within the perceptual state of the observer. Clearly, the metasemantic strategy for solving (GIC) brings much less metaphysical baggage.

^{xii} (Chalmers 2012) explores different versions of the basic accessibility relation at the heart of the proposal. The constraint we propose here corresponds to Chalmers' 'Conditional Scrutability' in its non-generalized form. Christopher Peacocke defends a similar accessibility constraint for concepts in general (Peacocke 1992) and for moral terms in particular (Peacocke 2004, chap. 7). Reference assignments must be justifiable on the basis of a core inferential role that constitutes the possession conditions for the concept. In the moral case, for instance, Peacocke appeals to a restricted set of a priori moral principles implicit mastery of which are necessary for possessing moral concepts and from which all other moral principles can be derived. But (IA) is not committed to this foundationalist structure of normative epistemology: e.g. you may need to take into account your whole implicit 'folk theory' about the normative domain in order to reach ideal verdicts about the extension of normative terms in all possible worlds (Jackson and Pettit 1995). (IA) is also consistent with the claim that you must know about your actual empirical circumstances, including your social linguistic context, to identify the instantiation of normative properties (Schroeter and Schroeter in press b).

^{xiii} Chalmers also includes a complete description of the phenomenal properties in his base (Chalmers 2004). This addition is necessary if you believe that phenomenal properties are not reducible to physical properties. We'll ignore this complication here.

^{xiv} Ideal Accessibility is, of course, compatible with a great deal of actual fallibility. Knowledge of the relevant facts may even be beyond our actual cognitive capacities: given our actual cognitive limitations, it may be impossible for us to gain access to the relevant empirical facts, the amount of empirical information required for full information may be impossible for us to process, the reasoning involved may be beyond our actual cognitive powers, etc. See Chalmers (2012, chap. 3) for a discussion of the relevant idealizations.

^{xv} The fact that we invoked Chalmers and Jackson in formulating (IA) may be misleading in this respect. In addition to embracing a metasemantic constraint like (IA), they also advocate a number of other theses that are extraneous to our interests here: e.g. they hold (i) that the semantic content of all terms involves two distinct intensions mapping possible worlds to extensions, (ii) that these intensions are generated solely on the basis of the subject's current internal states, (iii) that one of these intensions captures application conditionals that are knowable a priori, (iii) that any change in either of the intensions constitutes a change in meaning. These further claims have proven highly controversial in the philosophy of mind and language (see Schroeter (2017) for an overview). However, endorsing (IA) together with Determinacy does not commit one to any of them.

^{xvi} See Balaguer (1998). The mathematical and normative cases differ in a number of important respects, however. First, normative predicates apply to ordinary empirical objects, whereas mathematics posits mathematical objects and properties whose relationship to empirical facts is unclear. This makes characterizing the instantiation conditions much more straightforward in the case of normative properties than in the case of mathematical ones. Our solution to (GIC) for Autonomous Normative Naturalism depends on the fact that normative predicates have empirical instantiation conditions, so it cannot be straightforwardly generalized to mathematical Platonism. Second, Balaguer emphasizes that the only constraint on mathematical truth is internal formal consistency. In the normative domain, however, it is uncontroversial that there must be further constraints on normative truth than mere internal consistency that can single out determinate empirical applicability conditions for normative predicates. For a related worry about whether Balaguer's position can capture mathematical objectivity see Schechter (2010).

^{xvii} Recall that we are interested in the question of reliability insofar as it provides a good benchmark for progress in bridging the chasm highlighted by Benacerraf. We can thus afford to remain non-committal about how exactly reliability is relevant to knowledge. A near consensus view in epistemology is that knowledge proceeds from reliable processes or dispositions. But see Turri (2015) for arguments that reliability is not strictly necessary for knowledge. For an overview of different takes on the relevance of reliability for knowledge, see Goldman and Beddor (2015).

^{xviii} Focusing on subjects' reliability about the *instantiation of properties* helps sidestep a puzzle highlighted by Justin Clarke-Doane (2012; 2015). Clarke-Doane focuses on the reliability of beliefs about necessary moral truths such as 'torturing toddlers for fun is morally wrong'. Since there is no possible world in which such beliefs are false, we cannot appeal to ordinary counterfactuals in assessing the reliability of the dispositions generating such beliefs. But no such puzzle arises for Counterfactual Correlation, since it focuses on contingent normative truths like 'this action is morally wrong' (pointing to an instance of toddler torture). In testing the reliability of the dispositions generating this belief, we hold fixed the metaphysically necessary instantiation conditions for normative properties, and simply focus on whether variation in their instantiation affect the subject's beliefs about their instantiation. This exactly parallels our test for counterfactual reliability in the Nepalese village case: we hold fixed the instantiation conditions for *being a house*, and ask whether variations in the existence of houses in the village will affect the subject's beliefs about their existence. We don't ask whether dispositions generating the belief that buildings designed and used for human habitation are houses are counterfactually sensitive to variations in necessary truths about what it takes to be a house.

^{xix} Without (IA), (CGC) rests on a metasemantic mystery. (CCC) simply *stipulates* that *if* your N-criteria are ones you'd endorse at the ideal limit of informed reflection, *then* those criteria are reliable. But why should the world cooperate in this way, making our ideal verdicts reliable? In the absence of a metasemantic link between one's ideal understanding of 'N' and its reference, it's hard to see how one can avoid appealing to brute luck somewhere in one's overall philosophical account of the domain. An adequate answer to (GIC) seeks to avoid this sort of mystery.

^{xx} Two-dimensional semantics can be used to mark this distinction between reference-fixing conditions and the conditions necessary for the existence of the reference (the instantiation conditions): see Kaplan (1989) for an application to indexicals, Jackson (1998), Chalmers (2004) for applications to the semantic content of names and natural kind terms, and Stalnaker (1978; 1999) for applications to the content of speech acts and partial understanding of referential semantic content.

^{xxi} The distinction between metaphysical and metasemantic mind-independence has not always been recognized. Theorists like Michael Dummett (1978; 1991) and Hilary Putnam (1981; 1983), for instance, have sought to characterize metaphysical realism in metasemantic terms – roughly as the thesis that a subject’s ideal understanding of what it takes for a sentence to be true could turn out to be mistaken. However, we think metaphysical realists like David Lewis (1984) and Michael Devitt (1991) have provided decisive reasons to reject this characterization of their metaphysical position. Devitt sums up the point well:

“What has truth to do with Realism? On the face of it, nothing at all. Indeed, *Realism says nothing semantic at all* beyond [...] making the negative point that our semantic capacities do *not* constitute the world. (Devitt 1991, 39; see also Lewis 1984, 231-2)

^{xxii} Not all versions of Autonomous Normative Realism aim to vindicate the strong sense of normative objectivity advocated by Enoch. Proponents of what we earlier called the ‘anti-metaphysical’ approach to (ANR), for instance, may allow that different communities could pick out different properties with their normative terms (see for instance Scanlon (2014), and Enoch (2011, 121-7), for a critique). So not all proponents of (ANR) may endorse this second objection to our answer to (GIC). On the question of normative objectivity, convergence, and being hostage to psychological fortune, see also Setiya (2012, 118-127).

^{xxiii} One might wonder whether we should accept (SU) as a constraint on an adequate semantics for (ANR). It’s important to note that (SU) can’t be motivated by appeal to intuitions about some generic disagreement or conflict between communities using buck-stopping concepts. (SU) assumes that the kind of disagreement in question is *semantic*: conditions for competence with the relevant concepts guarantee that speakers in the different communities co-refer, regardless of variations in their social, psychological and empirical circumstances. This is a controversial and theoretically loaded semantic thesis. The mere fact that we’d feel *some sort of* disagreement with foreign language speakers who apply their buck-stopping predicate differently does not establish the theoretical claim that our terms are semantically guaranteed to co-refer (Plunkett and Sundell 2013; Väyrynen 2013: 66-9; Dowell 2015).

^{xxiv} In particular, van Roojen, Dunaway and McPherson, and Williams all hope to show how normative realists can accommodate intuitions about shared meanings in Moral Twin Earth cases, where different individuals or communities come to divergent verdicts about the applicability of normative concepts.

^{xxv} This is basically the position Lewis sketches in his response to Putnam (Lewis 1984). Several theorists have pointed out that Lewis’s own considered view is more complex, drawing on his account of radical interpretation of the subject’s non-ideal cognitive states in her current environment (Williams Ms.).

^{xxvi} By the same token, the more weight accorded to eligibility over charity, the better-placed reference magnetism will be to vindicate the stability of reference despite rationally irreconcilable normative disagreement. Many normative realists will see this potential for greater fallibility and increased referential stability as an advantage of for (RM) over (IA).

^{xxvii} This paper was inspired in part by discussions with David Enoch, Justin Clarke-Doane, and Bill FitzPatrick. We’d like to thank them for their generous engagement on the topic. We’d also like to thank Derek Baker, Colin Marshall, Tristram McPherson, Greg Restall, Kieran Setiya, several anonymous journal referees, and audiences at the University of Melbourne and the Australian National University for very helpful feedback on earlier drafts of this paper.