Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

Bad News for Ardent Normative Realists?

LAURA SCHROETER
The University of Melbourne
&
FRANÇOIS SCHROETER
The University of Melbourne

According to Ardent Normative Realists, reality favors certain ways of valuing and acting. Matti Eklund has recently argued that Ardent Normative Realists are committed to Referential Normativity, i.e., the thesis that the action-guiding and motivational roles associated with normative predicates determine their reference. In this paper, we argue that Referential Normativity should be rejected.

In metaethical debates, 'normative realism' is often used to pick out a minimal set of commitments: (i) normative statements are truth-apt, (ii) some atomic normative statements are true, and (iii) the properties or relations picked out by normative terms are metaphysically mind-independent. This Minimalist Realism secures the truth and objectivity of normative claims.

However, some metaethicists have sought to defend a more robust and metaphysically committed brand of normative realism (see, for instance, non-naturalists such as FitzPatrick 2008; Enoch 2011; and naturalists like Dunaway & McPherson 2016). Following Matti Eklund, we'll call such a position "Ardent Normative Realism". Roughly, the idea is that "reality itself favors certain ways of valuing and acting" (2017: 1). Two communities may rely on different evaluative frameworks in assessing their actions: that is, they may have distinct criteria for determining which actions instantiate the normative properties they attribute to actions. Yet according to Ardent Realism, some ways of valuing actions are privileged "from the point of view of reality itself" (Eklund 2017: vii). The Ardent Realist's suggestion, then, is that if two communities endorse

Contact: Laura Schroeter <laura.schroeter@unimelb.edu.au>
François Schroeter <fschro@unimelb.edu.au>

https://doi.org/10.3998/ergo.1135

incompatible normative frameworks, they can't both be getting it right on normative matters: at most one of them can "limn" the normative structure of reality (Eklund 2017: 22).

In this paper we focus on Eklund's claim that Ardent Normative Realists are committed to *Referential Normativity*, a metasemantic thesis about how the reference of normative terms is fixed by the distinctive normative conceptual role they play in practical reasoning. The bulk of the paper will be a critical examination of this thesis. As we'll see, Referential Normativity has played an important role in metaethical debates over disagreement and objectivity, so the thesis is of general interest independently of its role in supporting Ardent Normative Realism. We'll argue, however, that Referential Normativity should be rejected. If Eklund is right that Ardent Normative Realism is committed to Referential Normativity, then this is bad news for Ardent Normative Realism.

1. Referential Normativity

Imagine a foreign linguistic community that associates their term 'all-told right' with the same normative role we English speakers associate with our term 'all-told right'.[1] In particular, the normative judgments they express with this term play the same buck-stopping role in deliberation and the guidance of action as the judgments we express with the same term. Could it be that this community picks out a different property as the reference of their term 'all-told right' from the property picked out by our own use of that term? Following Eklund, we'll call this scenario *Alternative Reference*.

If Alternative Reference is possible, a judgment of the form 'a is right' about a particular action a might be true when made by members of the foreign linguistic community, but false when made by members of our own linguistic community. So strictly speaking, the two communities are talking past each other when they use this term: they're talking about distinct properties. But according to the Ardent Realist, at most one of the two communities can "limn" the normative structure of reality.

To bring out a challenge for this position, Eklund focuses on a *Further Question* that he takes to naturally arise for the Ardent Realist once we raise the possibility of Alternative Reference (2017: 22–32). Which community's use of the term 'right' succeeds in limning the normative structure of reality? Is it our own use, that of the foreign community, or are both communities off target? Eklund is interested in how Ardent Realists within the two communities should express this Further Question. Obviously, asking which actions are really the 'right' ones

[1] From now on, for ease of exposition, we'll use 'right' instead of 'all-told right'.

won't do. Suppose we English speakers use our normative terms to attempt to ask the Further Question. The right actions will then simply be the ones to which our term 'right' applies. We'll have thus failed to express the Further Question, the point of which is precisely to challenge that assumption. But what alternative vocabulary could do this job?

According to Eklund, there is no easy answer to that question and the Ardent Realist is faced with an *Ineffability* worry. While she insists that the possibility of Alternative Reference raises a Further Question, the Ardent Realist is in the uncomfortable position of being unable to express that Further Question (2017: 18–19, 23–29).

Now, Ardent Realists can avoid the Ineffability worry if they reject Alternative Reference. This can be done by embracing a substantive metasemantic thesis, *Referential Normativity*:

> (RN): two predicates or concepts conventionally associated with the same normative role are thereby determined to have the same reference. (Eklund 2017: 10)

Let's return to our initial scenario. Given that both we English speakers and the foreign linguistic community associate 'right' with the same normative role, Referential Normativity entails that the two communities pick out the same property. The two communities can thus formulate the Further Question simply by using their respective terms 'right'.

According to Eklund, the Ardent Realist has "no other reasonable choice" but to appeal to Referential Normativity, which he takes to be "the only possible viable way out" for the Ardent Realist—given the need to avoid the Ineffability worry (2017: 10–13). It is worth noting that Eklund himself, although attracted to Ardent Realism, wishes to remain neutral about the truth of Referential Normativity (2017: 12).

Other metaethicists have explicitly embraced metasemantic approaches in the spirit of Referential Normativity (see for instance Wedgwood 2001; Gibbard 2003; Enoch 2011; Williams 2018). It's not hard to see why Referential Normativity should have a broad appeal for metaethicists generally, and not simply to help Ardent Realists solve the Ineffability worry. What Referential Normativity promises is to sidestep all substantive disagreements in determining the semantic value of normative judgments. As long as different individuals or communities associate 'right' with the standard normative role, they are guaranteed to pick out the very same property—irrespective of how radically they may disagree about the nature of that property or about which actions instantiate it. This means modern humans and intelligent spiders would be able to directly logically disagree about normative matters, provided they had terms associated with the same normative role. Major metasemantic worries about the possibility of genuine normative disagreement could then be put to rest. But is Referential Normativity, with all its promises, too good to be true?

2. Two Constraints on Interpretation

The problem with RN, we'll argue, is that the normative role by itself cannot justify the ascription of any specific property as the reference of a normative predicate. To show why, we highlight two core theoretical roles played by semantic contents in the philosophy of mind and language. These theoretical roles generate important constraints on semantic interpretation: any plausible semantic interpretation must be able to play these roles. However, we'll argue that RN posits semantic contents that cannot fulfill these core theoretical roles. If this is right, RN is not a meta-semantically viable position.

The goal of semantic interpretation is to single out a specific semantic content for a given expression from amongst all possible candidate contents. In the case of 'right', for instance, an Ardent Realist expects a correct semantic interpretation to rule out candidate semantic assignments such as: (i) the term is meaningless, (ii) it expresses only a non-cognitive attitude or inferential role on the part of the speaker, (iii) the term is an ad hoc tool for partitioning locally salient cases without attributing any specific property, (iv) the term attributes different properties depending on the context of use and/or the context of interpretation, or (v) the term attributes a property that is not (or cannot be) actually instantiated. Instead, the Ardent Realist expects a correct interpretation to assign a specific property as the reference of all competent uses of 'right'—a property that captures perspective-independent normative reality. This property, moreover, will have empirical instantiation conditions that specify precisely which actions are right for which agents across all possible worlds.[2]

According to RN, any two predicates that are conventionally associated with the same normative role are thereby determined to have the same reference. In

[2] In other words, the property picked out by 'right' is instantiated by physical objects and events. In contrast, the property picked out by 'prime number' has no empirical instantiation conditions, since it is not instantiated by physical things. We are also assuming here that normative properties have their empirical instantiations conditions essentially: differences in the instantiation conditions entail differences in the property. We take these assumptions to be uncontroversial in the present context. Even metaethical non-naturalists agree that normative properties are empirically instantiated and globally supervene on the empirical properties of actions. What non-naturalists claim is that normative properties cannot be *identified* with their empirical instantiation conditions. For more details on the metasemantic task in metaethics, see Schroeter and Schroeter (2017). On the importance of metasemantics for the defense of metaethical positions and their overall plausibility, see Schroeter and Schroeter (2019).

other words, the correct methods for semantic interpretation must assign the same referential content for any predicate governed by this normative role. So semantic interpretation operating on normative role must suffice to ground:

(a) Stable referential purport: all competent uses of a thin normative predicate 'N' have the *semantic function* of attributing the very same property; and

(b) Reference-fixing: there is a *specific empirically instantiated property* that is attributed by any competent use of the normative predicate 'N'.

By itself, (a) rules out (i) semantic unintelligibility, (ii) purely expressive or inferentialist semantic content (Blackburn 1998; Gibbard 1990; 2003), (iii) semantic localism (Rayo 2013), and (iv) contextualist (Harman 1975; Dreier 1990; Finlay 2014) and relativist (Kölbel 2002; MacFarlane 2014) semantic contents. However, (a) is compatible with (v) fictionalism (Joyce 2001; Kalderon 2005) and error theories (Mackie 1977; Streumer 2017) as well as (vi) context-invariant realism. The reference-fixing requirement (b) rules out (v). In order to secure determinate and stable reference to a property, the correct methods for semantic interpretation must distinguish the property picked out from all other possible properties—including properties whose empirical instantiation may diverge only slightly in distant possible worlds or distinct properties that are necessarily co-extensive (if there are any). We will argue that correct methods of interpretation are not capable of securing either (a) or (b) solely on the basis of the normative role of thin normative terms.

Our basic worry about RN is that there is a mismatch between its highly restricted input into semantic interpretation (the motivational, emotional, and action-guiding roles of judgments involving a term) and the robust output of interpretation (a unique property with specific empirical instantiation conditions). The question is how such restricted input can ground such a rich semantic output. A proponent of RN needs to explain why this semantic assignment is warranted by plausible general principles for semantic interpretation. In particular, RN must avoid positing ad hoc interpretive principles specifically tailored to generate the desired verdict for the case at hand. The correct interpretive principles must generate plausible verdicts across the board for both normative and non-normative terms.[3]

[3] The requirement of uniform interpretive principles may seem to beg the question against Eklund. Eklund floats the 'metasemantically radical' suggestion that there may be two distinct types of reference-fixing relation, one for ordinary descriptive predicates, the other for normative predicates (Eklund 2017: 43). Eklund likens this position to dualism in metaphysics. However, the analogy isn't quite apt. Even if there are two distinct reference relations connecting predicates to the properties they attribute, we still need a general explanation of why referential content is

In Sections 3 to 5, we'll explain our worries about stable referential purport and reference fixing. But first we need get clearer about the nature of semantic interpretation. Semantic interpretation can be thought of as a function that takes as input empirical facts about a particular use of a term and yields as output a specific semantic content for that use. The empirical inputs into interpretation might include, for instance, the individual's most resilient inferential or recognitional dispositions to use a term, or an explicit stipulation of a semantic rule governing its use, or implicit presuppositions about felicitous or infelicitous uses. Inputs may also include facts about how an individual's current use of the term is related to wider aspects of their historical, social, and physical environment. The interpretation function then uses general rules to assign specific semantic contents to expressions on the basis of this empirical information about the use. The output of interpretation might assign a proposition (e.g., truth-conditions), a reference (e.g., an object, kind or property), a function (e.g., Kaplanian character), a truth-function (e.g., conjunction or negation) or quantifier (e.g., universal or existential), a conventional expression of an emotion or motivation (e.g., approval, intention to perform, disgust), and so on. In short, the interpretation function is a general method for assigning semantic contents to all possible expressions, which explains how *these* empirical inputs generate exactly *those* semantic outputs.

The precise nature of this interpretation function is hotly contested. Is the correct semantic assignment a direct reflection of the individual's current, most resilient dispositions to use the term? Or does the correct assignment sometimes depend on causal relations to objective features of the environment? How, if at all, does an individual's relations to her own past use of a term and to her linguistic community's use affect the semantic assignment? Do the methods for semantic interpretation require that we work from the bottom up, first assigning contents to particular terms, which then settle the contents of whole sentences? Or do they require us to work from the top down, assigning truth conditions to whole sentences first, which then fix the contents of individual terms? We won't seek to settle these questions here. Instead, our aim is to isolate more general constraints on any adequate theory of semantic interpretation.

Happily, there is widespread agreement about the core theoretical roles played by semantic contents. Our focus here is on the contents of expressions

warranted in the first place. Why should we assign a stable reference, rather than (e.g.) a purely expressive content or a context-variable content? Explaining why certain terms should be interpreted as having a referential function requires a general account of the explanatory role of assigning referential semantic contents. The correct interpretive principles must respect these general explanatory roles. Sketching the elements of these explanatory roles and the interpretive principles they ground is the task we take up in the text. For more details about our approach to referential purport, see Schroeter and Schroeter (2018).

in individual idiolects, where word meaning and thought content are closely aligned.[4] These contents play an indispensable role in predicting, explaining, and evaluating individual speakers' words and thoughts. In the philosophy of mind and language, it's common ground that an adequate method of semantic interpretation must generate contents that can play these theoretical roles. Let's unpack more carefully two central theoretical roles semantic contents play and the constraints these roles place on methods of interpretation.

Of course, not all psychological explanations involve semantic content ascriptions: the startle response or the effects of sleep deprivation do not hinge on the contents of words or thoughts. But it's uncontroversial that *content ascriptions play a central role in everyday psychological explanation*. For instance, Theresa May's decision to resubmit her Brexit plan to Parliament for the third time is explained by the specific contents of her prior beliefs and desires and the (subjectively accessible) logical and evidential relations linking those contents. This sort of content attribution lies at the heart of ordinary psychological explanation. An important constraint on an adequate method of semantic interpretation, then, is that the contents it ascribes must be suited to figuring in this sort of psychological explanation.

i. <u>Psychological Explanation Constraint</u>: Semantic contents must ground rationalizing explanations of an individual's reasoning and action.

As is standard in philosophy of mind and language, we call these content-based psychological explanations *rationalizing explanations*—that is, causal explanations which appeal to rational relations among thought contents, such as content identity, entailment, and evidential relations.

It is uncontroversial in the philosophy of mind and language that the role of content ascriptions in rationalizing explanation requires that the semantic contents attributed should (at least partly) *reflect the individual's actual patterns of understanding*. In order to serve in prediction and explanation, for instance, content ascriptions must reflect commonalities in individuals' dispositions to use concepts in categorization and inference, as well as their awareness of rational relations among thought contents, such as content identity, contradiction, entailment relations, and so on. For instance, if we want to explain how and why May's Brexit strategy changed over her time as PM, we must assume that her implicit understanding of which things would fall into the extension of terms

[4] We'll assume the reference of a term in an individual's idiolect is the same as the reference of the concept they express with that term. The core issues at stake in metaethics (e.g., reference, co-reference, and logical disagreement) arise at the level of concepts and idiolect meanings. So we can set aside the question of how idiolect meaning and public language meaning may diverge for present purposes.

like 'Brexit' or 'Europe' at any given time was (at least partly) accurate. If we were to say that May's use of these terms referred to the number three and polynomial equations respectively, our semantic interpretation would be useless in predicting and explaining May's actual reasoning—for those interpretations have no relation to May's actual cognitive dispositions to apply these terms and to use them in inference.

In addition, we must assume that (at least some) of May's words like 'Brexit' had stable semantic content over time—and that she was implicitly disposed to treat them as such in her reasoning. For instance, consider the following inference:

a. If I don't deliver Brexit, I will have failed. (formed in 2017)
b. I haven't been able to deliver Brexit. (formed in 2019)
c. Therefore, I have failed. (formed in 2019)

In order to recognize the validity of this line of reasoning, May must implicitly recognize that the semantic contents expressed by 'Brexit' in the two premises are non-accidentally the same. Non-accidental sameness of content is essential to validity, regardless of whether the contents are referential, context-dependent, indeterminate, or simply incoherent.

A second role for semantic content ascriptions is in the *evaluation of the truth* of an individual's assertions, beliefs, and implicit assumptions involving a particular concept.

ii. Normative assessment constraint: Semantic contents must determine truth-conditions that set standards for assessing the correctness of assertions, beliefs and cognitive dispositions.

This role of semantic contents in fixing truth conditions for thought and talk imposes a distinct, and partially competing constraint on a method for semantic interpretation. Obviously, not all beliefs are true. Theresa May's belief that she could deliver Brexit, for instance, turned out to be false; and when May ultimately abandoned that belief, the change in her overall state of understanding did not alter the semantic content of 'Brexit' in her idiolect. So the assignment of contents cannot vindicate *every* aspect of the thinker's current understanding associated with her words—some of her beliefs are likely to be false and some of her inferential dispositions mistaken. Indeed, according to our best common sense standards, individuals can be ignorant or mistaken even about the defining characteristics of the objects, kinds, properties or relations our words and thoughts pick out. Familiar externalist thought experiments bring this point home vividly: Oscar may not know how to distinguish the stuff he refers to as 'water' from superficially similar liquids (Putnam 1975), Bert may not realize that what he calls 'arthritis' is necessarily a disease of the joints (Burge 1979), and moral theorists can't figure out what precisely they are talking about with the term 'morally wrong'.

A central theoretical role for semantic contents, then, is to help distinguish between veridical beliefs and cognitive dispositions from non-veridical ones. For instance, consider a case of sustained scientific inquiry, in which Dalton, Thomson, Rutherford and Bohr all take themselves to be offering different theories about the nature and structure of atoms. Semantic interpretation of the contents of their words and thoughts sets standards we use to determine which of the proposed theories (if any) are true. The interpretive method also determines whether these theorists are all co-referring to the same kind of entity or whether they are talking past each other. An adequate method for semantic interpretation should assign semantic contents that set plausible normative standards both for evaluating the truth of specific beliefs and for the success of rational inquiry into the nature of the topic picked out.

These two constraints on interpretation—explanation and assessment—have been common ground in the philosophy of mind and language since Frege. Frege introduced a specific notion of semantic content, *sense*, in order to explain why individuals may reject true identity claims like 'Hesperus = Phosphorus' (Frege 1892). Fregean senses are descriptive reference-fixing criteria that are assigned in a way that reflects an individual's core criteria for identifying the reference of a word. Sameness or difference in core criteria associated with a name is supposed to explain why some identity claims will seem like trivial logical truths, while others seem to depend on contingent empirical facts. Since the assignment of senses reflects the individual's own cognitive dispositions, these semantic contents are suitable for rationalizing explanation of an individual's reasoning and action. At the same time, the Fregean promises to set plausible normative standards for assessing the truth or falsity of an individual's beliefs. Fregean senses determine the essential defining characteristics of the reference—not all empirical facts about it. So, although the individual's grasp of sense is incorrigible with respect to these defining characteristics, there is scope for ignorance and error about further empirical facts about the reference.

The problem highlighted by semantic externalists is that we often don't seem to have any infallible core criteria for identifying the reference of our words and thoughts. Indeed, settling our essential reference-fixing criteria in advance would implausibly restrict rational inquiry into the real nature of familiar objects, kinds and properties we seek to represent in thought (Putnam 1970; Burge 1986; Millikan 1984; 2017). Externalism has been motivated primarily by reflection on our normative standards for assessing the truth conditions for thoughts and the success conditions for rational inquiry. Our reflective judgments about the standards to which we hold individual thinkers rationally accountable do not seem to be grounded in any prior reference-fixing criteria in the way required by traditional internalist theories of semantic interpretation. So the motivation for externalism stems largely from the normative assessment constraint—from our reflective judgments about the appropriate standards for epistemic success. However, semantic externalists have not abandoned the explanatory constraint on interpretation. Externalists as diverse as Devitt (1981), Dretske (1988), Davidson (1973; 1987), Burge (1986), Williamson (2007) and Millikan (2017) all hold that content attributions figure in ordinary causal explanations of behavior—despite their disagreements over how exactly semantic contents are grounded in the historical, environmental, social and cognitive facts about individual thinkers.

The important general lesson for our purposes is that these two core constraints on content attribution—rationalizing explanation and normative assessment—are common ground among theorists of very different persuasions. These constraints pull in slightly different directions. The rationalizing explanation constraint requires theorists to assign semantic contents that (at least generally) *reflect an individual thinker's actual dispositions to use an expression*. But the normative assessment constraint on interpretation requires theorists to avoid making the individuals' actual dispositions infallible. More specifically, the constraint requires semantic contents to *set normative standards for evaluating the truth of an individual's words and thoughts*. Different theories of interpretation propose different ways of balancing the tension between these two roles.

Opinions about normative standards differ, of course. But in general the effort to meet the normative assessment constraint has led theorists to conclude that normatively acceptable content attributions must take into account external facts about the history, linguistic community, natural environment, and reflective practices historically linked to the individual's use of a particular term. If they are to play the relevant role in setting plausible normative standards of correctness, however, these external factors *cannot be wholly divorced from the individuals' understanding and use of a term*. Standards for epistemic success *must be grounded in the individual's prior practice*.

This point will be important in our discussion of RN, so it is worth spelling it out more carefully. Consider scientists' inquiry into the nature of atoms. Intuitively, scientists' theories of the nature of atoms seem to be empirically corrigible. (If we construed historical definitions of 'atom' as analytically true, we would have to say the term represented fictional entities rather than anything in the real world.) Dalton, for instance, took atoms to be indivisible particles of different mass whose recombinations explain all chemical reactions. But we don't just look to such theoretical definitions to decide what, if anything, his term 'atom' represented. Dalton's atomic theory was introduced as a way of explaining specific empirical observations about tin oxides, carbon dioxide, and other molecules. Looking back, we can point to the empirical facts about Dalton's explanatory aims, the actual history of experimentation he was focused on, the experimental apparatus involved, the fact that real substances were manipulated, and the fact that Dalton's theory captured the broad causal structures involved in the manipulated processes to support our semantic interpretation that Dalton was referring to real features of his actual environment—the features we now call 'atoms'. Thus *'external' facts about the explanatory project in which the term was actually used* seem to *justify our intuitive judgment about the real-world reference of Dalton's term*—despite misconceptions and ignorance involved in his theory. Moreover, an appeal to such external factors allows us to construe Dalton, Thomson, Rutherford, and Bohr as all referring to the same thing and progressively getting closer to the truth about the nature of that reference.

It's important to see that this interpretation is not just an arbitrary projection of our own contemporary understanding of the term 'atom' onto past uses of that term. Although normative standards for assessing the correctness of an individual's beliefs can violate some aspects of the individual's actual understanding of the term, they must still be justifiable on the basis of the individual's broader linguistic practices. To avoid positing arbitrary normative standards, a plausible interpretation of an individual's words must be *justifiable on the basis of their own understanding, background interests, and actual historical context in using a term*. For instance, Dalton's own goals and practices are what make it plausible that his term 'atom' referred to the actual chemical structures underlying the phenomena he was seeking to explain. Although Democritus held a similar formal theory, his explanatory aims were much less precise and there was no experimental history that could justify the conclusion that his use of 'atom' referred to *atoms*, rather than *molecules*, or *subatomic particles*. To interpret Democritus as univocally referring to atoms would be arbitrary—it simply cannot be justified on the basis of the empirical facts about Democritus's own practices with the term.

In sum, the two key roles of semantic content ascriptions—psychological explanation and normative assessment—place different constraints on an adequate method of interpretation. The explanatory constraint favors content ascriptions that closely track an individual's current classificatory and inferential dispositions, whereas the normative constraint favors content ascriptions that take into account broader 'external' facts about the individual's practice with a term. We have emphasized that these two constraints cannot radically diverge, even if they pull in slightly different directions. The normative assessment constraint cannot lead to normative standards that are divorced from the individual's own practices with a term. To avoid positing arbitrary normative standards, an interpretation must be grounded in these practices.

3. Normative Role and Interpretation

Let's now return to Referential Normativity. RN says that the normative role characteristic of the thinnest normative terms suffices for reference to a specific property. Interpretation must assign the same property as reference, regardless of the individual's other mental states, dispositions, or empirical circumstances. This means that the *only* input into semantic interpretation that's relevant to fixing the semantic content of a thin normative predicate is this normative role.

In a nutshell, our worry is that no plausible interpretation will be able to secure this result: the normative role does not provide rich enough constraints to single out a specific property as the semantic content picked out by normative terms. One difficulty in assessing RN is that the normative role itself is defined in an open-ended way, so that it's not clear what exactly could be included in that role. Of course, an account of the role should allow scope for theoretical variation—different theorists might favor different elaborations. But it's important to provide a substantive articulation of what is included (and excluded) in the conventional normative role (CNR) governing a predicate if we want to assess the claim that the conventional normative role, by itself, suffices to fix its reference. In this section, we'll provide an initial gloss on the normative role based on features highlighted in the literature. We'll then explain why the constraints on semantic interpretation don't seem to justify assigning a determinate reference—even a contextually variable one—on the basis of CNR alone. Indeed, the worry is that CNR does not even justify assigning a referential function to the predicates it governs. In the following section, we consider how CNR might be supplemented by further semantic conventions which might secure referential purport without violating the spirit of referential normativity.

Eklund understands normative roles as a characteristic role the predicate or corresponding concept conventionally plays in agents' practical deliberation, motivation, emotional response, and action (2017: 38). To illustrate the kind of role he has in mind, he appeals to Horgan and Timmons's Moral Twin Earth scenario:

> (H&T) Moral Twin Earthlings are normally disposed to act in certain ways corresponding to judgments about what is "good" and "right"; they normally take considerations about what is "good" and "right" to be especially important, even of overriding importance in most cases, in deciding what to do, and so on. (Horgan & Timmons 1992: 188; cited in Eklund 2017: 38)

Although Eklund remains agnostic about the precise character of the normative role of thin normative predicates like 'right', it is crucial to his argument that normative roles *exclude* any assumptions about empirical properties characteristic of right actions. On Eklund's understanding of the normative role, then, there could be individuals whose term, 'thgir', is governed by precisely the same normative role as our term 'right' even though their ultimate epistemic criteria for classifying acts as 'thgir' requires that they be acts of wanton cruelty. Despite this radical divergence in *epistemic role* played by these two terms, according to RN their similarity in *normative role* suffices to ensure that they have the same semantic content—both predicates attribute the same property.

A preliminary question we need to ask is why the normative role should be given overriding weight in determining the correct semantic interpretation of normative terms. When we assign semantic contents to a person's words, why should the normative role a term plays in an individual's mental economy always trump its epistemic role? As a general interpretive principle, this would be hard to justify. For instance, imagine a single-minded glutton, whose life is unreflectively and monomaniacally centered on the unhealthy pursuit of his favorite type of fast food, pizza. So the glutton's predicate 'is pizza' consistently plays a central deliberative, motivational, emotional, and action-guiding role of favoring the pursuit of things judged to fall into its extension (roughly, flat bread covered with tomato sauce, cheese and other savory toppings). And yet surely an adequate interpretation function would not treat this action-guiding role as an overriding factor in determining the semantic content of 'is pizza' in the glutton's idiolect. So why should the interpretation function treat the normative role of the glutton's terms 'is right' and 'is good' any differently? Why should the deliberative, motivational and action-guiding roles of these terms be an overriding factor in determining their semantic contents?

The answer cannot come from the general interpretation function itself. Instead, it must be grounded in empirical facts about the individual thinker's own mental economy. For instance, the glutton is probably disposed to treat the *epistemic role* of his predicate 'is pizza' as having overriding importance over its current *normative role*. So if his motivational priorities were to change, he would still call cheese-covered flat bread 'pizza'. In contrast, a denizen of Moral Twin Earth may be disposed to treat the *normative role* of her term 'is right' as strictly more important than her *epistemic criteria* for classifying actions as 'right'. So if her motivational priorities were to change, there would be a corresponding change in which actions she was disposed to classify as 'right'. This divergence in psychological dispositions could provide compelling empirical grounds for semantic interpretation to assign different contents to the glutton's predicate 'is pizza' and the Moral Twin Earther's predicate 'is right'. It's controversial, however, whether the individual's own psychological dispositions are decisive, and if they are just what form those dispositions must take.[5]

Fortunately, however, we can follow Eklund in abstracting from these disputes if we shift our attention to stipulative semantic conventions. Eklund is primarily interested in exploring whether there could *in principle* be a predicate that vindicates RN. So the cleanest case for his project is to focus on a predicate introduced via a stipulative definition. The conventional semantic rules governing a normative predicate governed by RN can be restricted to rules for the semantically appropriate consequences of normative judgments within an individual's mental economy.

For instance, we might stipulate a conventional semantic rule governing applications of a predicate 'N':

> (Conventional Normative Role: CNR) One should accept a judgment of the form 'a is N' only if one has (or goes on to have) a specific pattern of deliberation, motivation, choice, and emotional responses towards the action picked out by 'a'.

We can further stipulate that CNR is the *only* semantic rule governing the use of the predicate 'N'. In particular, there are no semantic rules governing the empirical conditions under which 'N' is correctly applied to actions. So competent users of the term are free to rely on any empirical criteria for applying 'N'—or they are free to apply the term at random with no stable criteria at all. With this stipulated semantic rule CNR in place, it's clear that when assigning a semantic content to 'N', the interpretation function should *not* pay attention to an individual user's current epistemic dispositions to apply the predicate to possible cases. The only factor that's relevant to the semantic interpretation is the stipulated semantic rule, CNR.

So what is the semantic content of 'N'? As we noted in the previous section, any acceptable interpretation function should assign contents that can play two core theoretical roles—*rationalizing explanation* and *normative assessment*. If we assume that CNR is the only semantic rule governing a competent speaker's use of a term 'N', what sort of semantic assignments can fulfill these two roles?

First, consider rationalizing explanation. A plausible interpretation function should assign semantic contents that help us predict and explain a competent individual's use of the term in reasoning and action. The more closely the

[5] There are tricky interpretive questions about precisely which psychological dispositions determine the relative weight an individual accords to different conceptual roles. Once we settle this general interpretive question, it becomes an empirical question whether any individual's use of 'right' actually accords overriding weight to a specific normative role.

attributed semantic contents mirror the cognitive dispositions guiding individuals' use of a term, the more reliable our content attributions will be in prediction and explanation. In the case of a predicate like 'N', which is governed by stipulative semantic rules, the core cognitive dispositions shared by all competent speakers will be those that are explicitly laid down by the semantic rules. So one interpretation of 'N' that would ground rationalizing explanations would be to assign it a purely *expressive content*:

> (Exp) 'a is N' serves to conventionally express a speaker's commitment to having specific patterns of deliberation, motivation, choice, emotional reaction with respect to the action picked out by 'a': {Da, Ma, Ca, Ea}.

Attributing Exp as the semantic content of 'N' would precisely mirror the semantic rule CNR accepted by competent users of the predicate. As a consequence, Exp allows us to reliably predict and explain their reasoning and discern commonalities in the reasoning patterns of different speakers. It would also allow us to keep track of subjectively accessible rational relations among an individual's thoughts. A sadist might exclusively apply 'N' to acts of wanton cruelty, but her predicates, 'is N' and 'is an act of wanton cruelty', are logically unrelated. Moreover, should the sadist's sadistic sensibilities change, this may alter her dispositions to apply 'is N' to cruel acts without altering her acceptance of CNR. So the expressivist content Exp would be better suited to predicting and explaining the sadist's actions and reasoning than a referential interpretation, which could depart in radical ways from the sadist's own dispositions to apply the predicate.

Let's now consider normative assessment. An interpretive function should assign semantic contents that set plausible standards for assessing when the individual's beliefs and utterances are correct. Whereas rationalizing explanation favors interpretations that closely mimic a subject's actual understanding, normative assessment of correctness favors interpretations that may depart in significant ways from the competent subject's current understanding of a term. As externalists pointed out, plausible normative standards for assessing the truth of an individual's words and thoughts may depend in part on facts beyond their ken, such as their actual historical, social, and physical context.

However, it is hard to see how an appeal to such external factors could help to vindicate a referential interpretation of 'N'. As we have emphasized at length in the previous section, a plausible semantic assignment should not set arbitrary normative standards for epistemic correctness: normative standards of assessment must ultimately be grounded in the individual's prior practice with a term. By hypothesis, the *only* aspect of a competent individual's use, understanding, and broader linguistic practice that's relevant to interpreting 'is N' is the stipulated normative role, CNR. But CNR only constrains which internal motivational and conative states can be combined with attributions of the predicate. By design, CNR is perfectly consistent with an individual relying on *any* empirical criteria for classifying actions as falling into the extension of the term—or with no consistent criteria at all. The problem is that there is nothing in this stipulated semantic rule that can justify one specific referential assignment over any other. Why should we hold thinkers normatively accountable to one specific referential standard of correctness rather than another? Privileging a particular referential assignment would be arbitrary from the point of view of the stipulated semantic rules. So plausible standards of normative assessment cannot ground any determinate referential assignment.

Neither rationalizing explanation nor normative assessment can justify attributing reference to any specific property on the basis of CNR. So CNR by itself fails to ground a determinate reference.

Indeed, there is reason to doubt that the term *has referential function* in the first place if it is in fact impossible to secure a specific reference. As we noted above, from the point of view of rationalizing explanation CNR seems to support an expressivist semantic interpretation (Exp) over a referential interpretation of the content of 'is N'. And from the point of view of normative assessment, it's hard to see how it makes sense to hold thinkers normatively accountable to a referential standard of correctness to which they have no rational access. This is not to say that error theories are never justified. But error-theoretic interpretations must be non-arbitrary. This means that to make the error-theoretic interpretation stick, the referential purport must be firmly anchored in the individual's own understanding and use of the expression. Consider the predicate 'is a witch'. Suppose our semantic rules stipulate that this predicate is correctly applicable to an individual only if that individual can magically fly over a barn on a broomstick. Clearly this criterion does not pick out anyone in the actual world. And one might argue that the 'magic' requirement is hopelessly indeterminate and so does not suffice to single out any determinate property. Yet it's intuitively plausible that a predicate governed by this rule has a referential function—even if it fails miserably in fulfilling at that function. The semantic rules stipulate what an object must be like in order for the predicate to apply to it, so those rules will strike the user as prima facie picking out a property of the object. In contrast, CNR is unlikely to generate this sort of intuitive referential appearance. When a semantic rule like CNR merely constrains what the user must *desire* or *feel* when applying it to an object, the predicate seems to perform an expressive function, not a referential one.

In sum, we suggest that neither of the two core theoretical roles for semantic contents—rationalizing explanation and normative assessment—supports a referential interpretation of words governed by CNR. If CNR is the only semantic convention governing the predicate 'N', it is hard to see why a referential interpretation would be more plausible than an expressive interpretation.

4. Securing Stable Referential Purport

In response, a proponent of RN could enrich the conventions governing thin normative terms. Perhaps 'is N' is governed by further conventions in addition to CNR, which ensure that the interpretation function assigns a referential content, rather than a purely expressive one. Whatever these further conventions are, they will presumably govern other paradigmatically referential terms, such as names, natural kind terms, artifact terms, and ordinary descriptive predicates. Of course, not all such terms actually succeed in singling out a reference. But terms like 'Zeus', 'zombie', or 'to hex' at least *purport* to stably represent specific features of the world. Something about our understanding and use of these terms seems to mandate a stable object, kind, or property as the semantic assignment. What the proponent of RN needs, then, is some explanation of what facts about the use of an expression 'N' would favor a stable referential assignment—what facts constitute an expression's *stable referential purport*? If we had such an account in hand, we could use it to supplement the conventions governing 'N' so as to favor a referential assignment for a purely normative term.

So what semantic conventions can we add to CNR would suffice to ensure referential purport? It's important, if we want to preserve Referential Normativity, that we *not* add any semantic conventions governing which empirical features of actions make 'is N' correctly applicable. Referential Normativity, after all, was introduced precisely in order to secure co-reference despite radical incompatibility in empirical application criteria.

One traditional strategy is to cite the predicate's role in grounding (what appear to be) standard *logical relations* among thought contents, such as inconsistency, validity, and entailment. For instance, the thin normative predicate 'is wrong' figures in valid argument forms like the following:

a. 'Stealing is wrong.'
b. 'If stealing is wrong, then getting your little brother to steal is wrong.'
c. 'Getting your little brother to steal is wrong.'

Entertaining logically complex judgments and patterns of reasoning is a core aspect of our use of normative predicates, just as it is for ordinary descriptive predicates. So we might add a further Logical Relation condition to the stipulations governing even the thinnest normative predicates:

(LR) It is semantically correct to treat the predicate 'is N' as embedding under logical operators and as grounding logical inference patterns in the same manner as ordinary descriptive predicates.

Prima facie, it's hard to see how a purely expressive interpretation of the normative predicate 'is wrong' could be embedded in the antecedent of the conditional in b. A purely expressive interpretation also seems ill-suited to explaining the strict logical incompatibility of accepting the premises while rejecting the conclusion. So LR may seem to count in favor a semantic interpretation that treats the predicate 'is N' as having the semantic function of attributing a stable reference.

However, LR is not decisive in favoring the stable referential purport of thin normative predicates—for there are alternative explanations of the logical role of normative predicates. Expressivists, for instance, have proposed a number of strategies for arguing that LR doesn't favor a referential interpretation of 'is N' over an expressive interpretation.[6] We take the jury to be out on the question of whether an expressivist semantic assignment for 'is N' is consistent with LR. Moreover, LR is clearly compatible with inferentialist/pragmatist semantic interpretations, which can simply build LR into the inferential role assigned as the semantic content of a normative predicate.[7] So LR does not decisively rule out non-representational interpretations of normative terms. In addition, LR is consistent with many other interpretations that stop short of stable referential purport. Contextualists seek to account for LR in terms of referential commitments within a conversational context, while relativists treat logical relations as structural commitments built into the conventional rules of use for a predicate. Finally, semantic localists like (Rayo 2013) resemble contextualists in treating logical relations as relative to a context, but deny that normative terms pick out a determinate property even within a given context.

A different suggestion that we find attractive is to ground referential purport in an expression's *stable classificatory role* over time and between individuals. Roughly, the idea is that certain ways of accumulating and managing empirical information constitute ways of *keeping track of empirical features of the world* in thought and talk. When an expression plays this classificatory role, it's plausible that semantic interpretation should favor assigning a referential content.[8]

[6] For an excellent overview of the debates, see van Roojen (2018), particularly section 4 and the supplementary document, 'Embedding Problem Response Strategies'. See also Baker and Woods (2015) and Pérez Carballo (2015) for defenses of expressivism that ground the logical properties of sentences in purely formal features of an expression's role.

[7] Prominent proponents of this approach to semantic contents include Brandom (1994), and Price (2011). For applications of this approach to thin normative terms, see Chrisman (2012; 2015) and Gert (2018).

[8] For more details on this approach to referential purport, see Schroeter and Schroeter (2018). Ruth Millikan explains how concepts acquire their reference along similar lines, but her approach

To get a better understanding of the conceptual roles distinctive of keeping track of (what the thinker takes to be) a stable empirical topic over time, let's consider paradigm referential terms like 'Gödel', 'gold', or 'golf'. Individual thinkers accumulate a body of standing attitudes and cognitive dispositions under these headings, which they automatically treat *as* pertaining to the same topic: for example, a man, Viennese, logician, lived in the 20th century, taught at Princeton, discovered the incompleteness theorem, wore glasses, looked like *that,* was called 'Gödel', etc. Such bundles of information ground an individual's classificatory dispositions to identify incoming information *as* pertaining to the same topic. This bundle of accumulated information is then available for inductive reasoning about a newly classified instance. Over time, the bundle of information associated with a given term will tend to grow as new information is gleaned from new empirical identifications. The more information stored in the bundle, however, the more scope there is for inconsistencies to arise between stored information and the information derived from new identifications. When such inconsistencies are detected, thinkers will normally seek to cull some attitudes or dispositions to eliminate incoherent empirical commitments. Thinkers may also engage in reflective theorizing in an effort to anticipate and resolve potential inconsistencies and hone more useful classificatory criteria. This theoretical reflection is often formulated in object-level terms: we may ask what it really takes to be Gödel, gold, or golf.

Similar patterns of information management govern linguistic coordination between individuals. We tend to hear others' use of paradigm referential terms like 'Gödel', 'gold', or 'golf' as obviously pertaining to the same topic we ourselves associate with those terms. Normally, an authoritative interlocutor's testimony about 'golf' becomes a direct source of information for inclusion in the bundle of information we associate with that term. Moreover, we're sensitive to inconsistency between the information our interlocutors associate with a term and our own prior commitments. When we identify a disagreement, we seek to restore coherence by culling some attitudes or dispositions within one of the disagreeing parties—preferably the ones that have a weaker justification. This effort at establishing coherence between interlocutors' associated attitudes can be aided by reflective theorizing about classificatory principles, which seeks to anticipate and resolve sources of incoherence and more precisely demarcated categories relevant to users of the term. Thus our interpersonal classificatory

is couched within a teleosemantic framework which we don't endorse. According to Millikan, our cognitive mechanisms naturally select for consistency in the bundle of empirical criteria associated with a particular concept (or 'unicept'). The historical process of selecting for consistency, she thinks, favors the survival of bundles of recognition criteria that all (reliably enough) target the same empirical feature (Millikan 2017: ch. 5). This is what warrants interpreting these concepts as having referential contents.

practices seem to be updated and managed with an eye towards coherence and usefulness in much the same ways as our intrapersonal practices.

Drawing on these observations, we can define the *Stable Classificatory Role* (SCR) of a term roughly as follows:

> (SCR) It is semantically correct to associate predicate 'is N' with an evolving bundle of attitudes and cognitive dispositions, which are (i) used to ground identification and induction, and (ii) monitored and revised to foster internal coherence and empirical usefulness in a context-neutral way, over time and between individuals.

Thin normative predicates like 'is right' seem to fit this pattern: we accumulate stable beliefs and implicit criteria about what it takes to be right, we engage in reflection and debate about particular cases, we formulate general theories about what makes an action right, and so on. The evolving bundle of attitudes and cognitive dispositions is relatively stable over time, transmitted via testimony, monitored for coherence, and refined through theoretical reflection. In short, our classificatory, epistemic, and theoretical practices with thin normative terms resemble our practices with paradigmatically descriptive terms.

So let's add both LR and SCR to our original stipulation that CNR governs the thin normative predicate 'is N'. Would this suffice to ensure that the interpretation function should seek to assign 'is N' a *stable referential content*? Prima facie, we think it would. But it's important to understand why.

Notice that this combination of constraints does not require the acceptance of any specific beliefs or classificatory dispositions involving 'is N'. A community of Moral Saints might use 'is N' governed by these stipulations, and they might tend to apply it almost exclusively to acts of selfless charity, while a community of Moral Perverts might use the term 'is N' governed by these same stipulations and be disposed to apply it almost exclusively to acts of wanton cruelty. SCR involves a commitment to formal patterns of reasoning—patterns that are characteristic of use and maintenance of coherent classificatory practices, suitable for keeping track of stable features of the world over time and between individuals. So adding SCR to the stipulations governing 'is N' ensures that all competent speakers will use 'is N' in classification and reasoning *as if* the term picked out a stable feature of the world whose nature is open to inquiry and debate. From the competent speaker's point of view, then, 'is N' will seem to have a stable reference. It's natural, then, to argue that the interpretation function should favor a stable referential interpretation other things being equal. This would ground the semantic interpretation in the practices and commitments of the individual thinker, vindicating the subjective appearance of referential purport.

It's plausible, then, that a predicate governed by the semantic conventions CNR + LR + SCR should be interpreted as having *stable referential purport*. This interpretation reflects central cognitive dispositions that help to *predict and explain* distinctive patterns in the individual's reasoning and behavior (as required by the Psychological Explanation Constraint). And given the centrality of these classificatory dispositions in structuring an individual's epistemic practices with a term, stable referential purport sets a non-arbitrary standard for assessing the *correctness* of the individual's use of the predicate (as required by the Normative Assessment Constraint). Given that normative predicate 'is N' shares the broad logical and classificatory roles characteristic of paradigm referential terms, there is strong reason to take the predicate to have stable referential purport.

However, it remains to be seen whether the conventions specified by CNR + LR + SCR will suffice to single out a specific empirically instantiated property as the reference. If the conventions do *not* fix a stable reference, the interpretation function may simply assign an empty reference for the normative predicate 'is N'—yielding an error theory for this normative term. Alternatively, we may have reason to revise the attribution of stable referential purport in the light of the impossibility of fixing a reference. Perhaps some form of semantic inferentialism, contextualism, expressivism or relativism would provide a better overall fit with the Psychological Explanation and Normative Assessment constraints on semantic interpretation. But we won't need to adjudicate this issue here, since our concern is to evaluate Referential Normativity, which requires normative predicates to pick out stable reference.

In the previous section, we suggested that the normative role, CNR, does not suffice by itself to support a referential interpretation of 'N'. In this section, we've proposed two friendly amendments to support a stable referential interpretation of 'N'. The additional semantic conventions we have proposed place purely formal constraints on an individual's internal patterns of reasoning involving the term—so they can be accepted regardless of the individual's empirical criteria for applying that term. If the enriched set of linguistic conventions, CNR + LR + SCR suffice to fix a determinate reference, then we will have vindicated the fundamental aims of Eklund's Referential Normativity thesis.

5. Reference-Fixing

Let's turn now to the question of reference-fixing. Do CNR + LR + SCR suffice to ensure that a specific property will be singled out as the reference of 'is N'?

It's important to keep in mind that Referential Normativity requires that all competent users of the term will pick out the same property, regardless of how their other mental states and external circumstances may vary. For instance, Anne may be disposed to apply the predicate 'is N' exclusively to *acts of selfless charity*, Beth may apply it to *acts of wanton cruelty*, Chris may apply it to *satisfying one's current whims*, and Doug may apply it exclusively to *acts of eating pizza*. We may further assume that these competent speakers each lives in an isolated community where their interlocutors are in rough agreement about how to apply the predicate. As long as each of these individuals—and their communities—accepts the conventions laid down by CNR + LR + SCR, Referential Normativity says their respective uses of 'is N' are guaranteed to stably pick out the very same property as its reference.

So just which property is the reference of 'is N'? It's generally uncontroversial within metaethics that the applicability of thin normative terms supervenes on the ordinary descriptive properties of actions: two actions that are descriptively indiscernible are normatively indiscernible.[9] To fix ideas, let's suppose that 'is N' picks out the property of *maximizing happiness* as its reference. (The arguments we'll present generalize to other candidate interpretations.)[10] Now, our question is whether such an interpretation can play the characteristic theoretical roles of semantic contents, Psychological Explanation and Normative Assessment.

First, consider Psychological Explanation. If we interpret all three users of 'N' as attributing the property of *maximizing happiness*, will this referential assignment help us to predict and explain their reasoning and actions? We think not. This interpretation may help predict and explain Anne—for example, her charitable actions, the evidence she takes into account in classifying actions as 'N', and her reflective theorizing about what counts as 'N'. However, this referential assignment is less than useless in understanding the reasoning and behavior of Beth, Chris, and Doug. And this worry seems to arise no matter

[9] Supervenience is a standard assumption among normative naturalists and anti-naturalists alike. To deny supervenience would be to opt for a radical property dualism, according to which the normative status of an action does not depend on the totality of descriptive facts about that action. This position seems to undercut both the epistemic and action-guiding roles of normative properties. See McPherson (2019) for an overview of supervenience theses in ethics.

[10] Non-naturalists might worry about our generalization to their view, since they deny that normative properties can be identified with any naturalistic property like *maximizing happiness*. However, non-naturalists generally agree that an action has a normative property in virtue of the ordinary naturalistic properties on which it supervenes. It follows that there is a necessary correlation between the instantiation of normative properties and the (perhaps complex and heterogeneous) natural properties in virtue of which they are instantiated. For ease of exposition, our example here simply *identifies* a normative property with the (complex) natural property in virtue of which it is instantiated. However, it makes no difference to our argument whether normative properties are numerically identical to natural properties. They key point is that normative properties have their naturalistic instantiation conditions essentially. On the importance of necessary instantiation conditions for metasemantics, see above, Footnote 2.

which referential assignment we choose: insofar as a property helps predict and explain the psychology of one of our four individuals, it will fail for the others. Moreover, we can't simply split the difference between them: there simply is no common ground among all competent individuals' criteria for applying a thin normative term. Indeed, Referential Normativity was introduced to overcome this radical incompatibility in competent speakers' epistemic criteria: the proposal was that reference is assigned solely on the basis of the predicate's role within an individual's practical reasoning—and *not* on the basis of its epistemic role in identifying what falls into its extension. The problem we're highlighting, however, is that this very fact makes any specific referential assignment irrelevant to explaining the epistemic aspects of competent individuals' psychology. The empirical instantiation conditions of the property picked out is of no help in predicting and explaining the reasoning and actions of competent users of the term. An error theory or fictionalist semantics would be just as useful for the purposes of psychological explanation—if not more so.

Now, let's consider Normative Assessment. This is where, one might suppose, Referential Normativity earns its keep: a stable referential assignment will set appropriate normative standards for determining which classificatory judgments involving normative terms are correct (i.e., true). The suggestion is that Anne, Beth, Chris and Doug all ought to classify an action as 'N' just in case that action has a specific property, like *maximizes happiness*, in the relevant circumstances. The reason is that this referential assignment makes it appropriate for such judgments to play the action-guiding, deliberative, emotional roles set out by the semantic convention (NR)—as well as vindicating LR and CR. Indeed, *everyone* should treat maximizing happiness as the appropriate standard for guiding action—regardless of their idiosyncratic psychology or circumstances. This, we take it, is the central intuition motivating Referential Normativity.

The key question, however, is whether this univocal reference-fixing claim fits with the general constraints on semantic interpretation. As interpreters, we might be tempted to project our own empirical criteria for deciding which actions to perform onto others—using our criteria to set the standards of correctness for other users of terms that play a normative role. But, as we have argued in our discussion of the Normative Assessment Constraint, plausible semantic assignments must be non-arbitrary: they must be grounded in the individual speaker's own understanding, use, and practice with a term. And the correct interpretation must be generated by universal principles of interpretation. In the case of 'is N', the only relevant input into semantic interpretation are the stipulative rules laid down by CNR + LR + SCR. Prima facie, these rules cannot single out one or another empirically instantiated property as the reference of 'is N'.

To see this, let's consider how our four competent speakers might seek to interpret the others' use of 'is N'. They all agree, we may suppose, that 'is N' purports to have a stable referential content, but they have very different views about the nature of the property picked out. Doug, the pizza-lover, may be tempted to interpret Anne the altruist, Beth the sadist, and Chris the wanton as all co-referring with himself. And the others may agree. So far, so good. But exactly which property are they all picking out? No doubt Doug will insist it's the property of *maximizing pizza consumption*. But this referential assignment seems utterly arbitrary from the point of view of the others. Doug is simply projecting his own idiosyncratic ultimate goals in interpreting others' use of the term. And if Anne were to interpret Doug's as referring to *acts of selfless charity*, her interpretation would strike Doug as similarly arbitrary. These four individuals have such radically divergent classificatory practices associated with 'is N' that it is hard to see how one speaker's classificatory standards could set non-arbitrary epistemic standards for correcting the classificatory practices of any of the others.

The challenge for a proponent of Referential Normativity, then, is to provide a non-arbitrary way of singling out a specific property as the reference for all four of the individuals we've considered. This referential assignment must be justifiable solely on the basis of the internal conceptual roles, CNR + LR + SCR, and the interpretative methods must be justifiable on general grounds. In particular, the correct referential assignment should meet the Psychological Explanation and Normative Assessment constraints on the assignment of semantic content. We don't see how this challenge can be met. The normative standards that can be derived from these semantic roles are themselves purely formal and action-guiding. All of our sample individuals should treat the term 'is N' as entering into logical relations like validity and incompatibility. All of them can be criticized for failing to conform to certain patterns of instrumental reasoning, for succumbing to weakness of will, or for failing to have feelings of remorse when they do so. And all four can be criticized for failing to update and store new assumptions about 'is N', to apply these assumptions to new cases, or to monitor this body of assumptions for internal incoherence. These diverse practical and epistemic standards of assessment are directly grounded in the conceptual roles fixed by our stipulative definition of 'is N'. But without some further constraints on interpretation, it is hard to see how we could reasonably criticize our four speakers for failing to apply 'is N' to all and only cases of *maximizing utility*. There is simply nothing in the conventions, CNR + LR + SCR, that could justify applying this specific standard for normative assessment.

Our conclusion, then, is that the kind of formal, action-guiding conceptual roles cited by Eklund will not suffice to fix a determinate reference for thin normative terms. But this does not yet show that Referential Normativity is false. If Referential Normativity is to be vindicated, however, it must look for further constraints on semantic interpretation over and above the internal conceptual roles we have outlined so far.

6. Reference Magnets

Let's consider one last strategy for vindicating Referential Normativity. So far, we have assumed that the semantic stipulations, CNR + LR + SCR, are the only topic-specific constraints on semantic interpretation. Recently, however, a number of metaethicists have appealed to David Lewis's notion of *reference magnetism* to explain how different individuals' use of a normative predicate might co-refer, despite their divergent epistemic standards for identifying its extension (van Roojen 2006; Edwards 2013; Dunaway & McPherson 2016; Williams 2018).

David Lewis (1983; 1984) originally proposed reference magnetism as a response to a different worry about semantic interpretation—indeterminacy. Suppose the interpretation function seeks to assign semantic contents to an individual's words in such a way as to maximize the truth of the individual's total set of beliefs. As Hillary Putnam pointed out (1980; 1981), this interpretive method will lead to radical indeterminacy of reference: we can construct indefinitely many gerrymandered sets of referential candidates that do equally well in maximizing the overall truth of the individual's beliefs. Indeed, distinct gerrymandered interpretations can make *all* of an individual's beliefs true, provided that their beliefs involve no formal contradictions. So long as the empirical constraints on interpretation are confined to the conceptual roles played by terms in an individual's idiolect, the interpretation function will yield radically indeterminate referential assignments. In response, Lewis suggests we need to add further, mind-independent empirical constraints on interpretation, that function as *reference magnets* to secure determinate reference to specific properties.

What is the nature of the empirical constraints needed to solve Putnam's indeterminacy problem? And how can this reference magnetism help secure non-accidental co-reference for normative terms? Different accounts of reference magnetism have emerged in the metaethical literature.

A first account turns on a general appeal to the metaphysics of *eliteness*. Relative eliteness is supposed to be a mind-independent metaphysical fact about properties, which could be explained in terms of grounding, relative fundamentality, or degrees of naturalness, etc. According to Douglas Edwards (2013) and Billy Dunaway and Tristram McPherson (2016), the interpretation of any referential expression can be thought of as a two-step process, first assessing *closeness of fit* and then *degree of eligibility*. In the first step, the interpretation function restricts the class of eligible referential candidates to those properties whose empirical instantiation conditions overlap 'enough' with an individual's actual classificatory dispositions. In the second step, the interpretation function selects the property in that range which has the highest degree of eliteness as the reference of the predicate. This metaphysical approach to reference magnetism promises to resolve Putnam's radical indeterminacy worry by positing *extra empirical inputs into the interpretation function*. Semantic interpretation must take into account metaphysical facts about the relative eliteness of different properties, in addition to facts about the conceptual role a predicate plays in an individual's mental economy. In effect, metaphysical eliteness works as a tiebreaker among the range of eligible referential candidates, ruling out gerrymandered properties and favoring the most elite properties in the range.[11]

Edwards and Dunaway and McPherson argue that metaphysical eliteness can also explain why different speakers' use of normative predicates can co-refer, despite their divergent classificatory dispositions. Whenever there is a unique, highly elite property that falls within the range of eligible referential candidates for both speakers' classificatory dispositions for a normative predicate 'is N', the interpretation function must assign the same reference to both. In such cases, the metaphysical eliteness constraint ensures the two speakers non-accidentally co-refer—even when their ideal, fully informed classificatory dispositions would diverge.

However, this metaphysical version of reference magnetism cannot vindicate Eklund's Referential Normativity thesis. Nor was it designed to. Edwards and Dunaway and McPherson hope to show that non-accidental co-reference is compatible with *some* divergence among individuals' ultimate epistemic criteria for applying a term. Crucially, however, this account of co-reference requires that individuals' classificatory dispositions must be *similar enough* to ensure overlapping ranges of eligible referential candidates. And these two ranges must include *exactly one* maximally elite property (located in the overlap). This interpretive method, however, cannot ensure that individuals like Anne, Beth, Chris and Doug—with their radically divergent classificatory dispositions—pick out the same property with their use of 'is N'. There simply is *no significant overlap* in the range of properties that would provide a close enough fit to the classificatory

[11] This account is based on Lewis (1984). Lewis himself took semantic interpretation to be *holistic*: all the words in an individual's idiolect (and the concepts they express) must have their reference fixed simultaneously. So the interpretation function must trade off (i) fit with an individual's totality of linguistic dispositions against (ii) naturalness of the totality of referential assignments for words in their idiolect. However, proponents of reference magnets in metaethics do not necessarily accept this holism. They assign classificatory dispositions for particular predicates considered individually, and evaluate relative naturalness for that predicate independently of naturalness of the total set of referential assignments.

dispositions of all four agents. So it isn't possible for metaphysical eliteness to select the same property for them all (cf. Dunaway and McPherson 669–70). Metaphysical reference magnets cannot secure non-accidental co-reference purely on the basis of CNR + LR + SCR.

Moreover, there are independent reasons to doubt that metaphysical magnets can ground a plausible semantic interpretation of thin normative terms. As several theorists have pointed out, reference magnets that are strong enough to secure co-reference across ordinary disagreements about the extension of 'is right' will, by the same token, be strong enough to secure co-reference with theoretical terms like 'maximizes utility' or 'satisfies the categorical imperative', which intuitively should not co-refer (Sundell 2012; Schroeter & Schroeter 2013; Williams 2018).

A second approach to reference magnetism seeks to avoid such implausible referential assignments by positing *topic-specific magnetic constraints* on interpretation. Thus, Mark van Roojen (2006) suggests that reference magnets should be *discipline-relative*: the kinds of properties that are eligible referential candidates for physics may be different than those that are most eligible for biology or for ethics. The suggestion is that the conceptual role of a predicate is tied to a specific epistemic discipline, and specific disciplines determine their own proprietary standards of property eliteness.

It's not obvious how to understand the notion of *a discipline* here, but presumably it will be cashed out in terms of epistemic aspects of the conceptual roles governing particular predicates. Predicates that play a role in physics and biology, for instance, might be associated (by an individual or their community) with specific inductive practices and specific ways of theorizing. Similarly, thin terms like 'is N' would be governed by specific inductive and theorizing practices that are distinctive of a specific normative domain. We might model this by adding discipline-specific constraints to the generic classificatory role, SCR, yielding different epistemic roles for different disciplines: ER_P for physics, ER_B for biology, ER_E for ethics, and so on. The general idea, then, is that each of these distinctive epistemic roles determines its own ranking of the relative eliteness of properties: the property of *being composed of carbon*, for instance, is highly elite relative to the epistemic practices in physics but non-elite relative to the epistemic practices in ethics.

There are a number of worries one might have about this proposal. One worry is that the approach does not succeed in imposing a *new empirical constraint* on interpretation. The ranking of relative eliteness of properties seems to be based on epistemic considerations—that is, the epistemic norms governing different disciplines—which are grounded in the internal perspective and epistemic priorities of thinkers themselves. But then, relative eliteness is simply part of the psychological (and social) inputs into the interpretation function, and it does not constitute an independent metaphysical constraint on interpretation. In that case, van Roojen's approach does not seem to address Putnam's original indeterminacy worry.[12] Even if this worry can be overcome, however, the discipline-relative account of reference magnets seems ill-suited to vindicating Referential Normativity, which seeks to ensure co-reference purely on the basis of CNR + LR + SCR. The problem is that this approach grounds the interpretation of normative predicates in the individual's specific inductive and theoretical practices. To secure the same standards of relative eliteness for their use of the thin normative predicate 'is N', two speakers must associate that term with the same disciplinary epistemic standards, ER_e. But it is highly implausible that speakers like Anne, Beth, Chris and Doug share the very same epistemic standards governing induction and theorizing about whether to classify an action as 'N'. By hypothesis, each of these speakers (and their respective communities) is disposed to reflectively converge on different properties as the most eligible referential candidate for 'is N': *acts of selfless charity*, *acts of wanton cruelty*, *satisfying one's current whims*, and *acts of eating pizza*. It's hard to see how these divergent verdicts could be grounded in some shared standards for induction and theorizing. So van Roojen's proposal seems of little use in defending Referential Normativity.

In contrast, Robbie Williams (2018) seeks to vindicate Referential Normativity. His strategy is to ground topic-specific reference magnetism for thin normative predicates exclusively in their motivational and action-guiding role. Unlike the other proponents of reference magnets in metaethics, Williams takes Lewis's account of radical interpretation as his model (Lewis 1974), rather than his response to Putnam's radical indeterminacy objection (Lewis 1984).[13] On Williams' account, the norms of rationality involved in radical interpretation are what make some referential candidates more eligible than others. The view hinges on the idea that radical interpretation cannot rely on purely *formal* or *structural* rationality norms if it is to determinately fix reference. We cannot, for instance, appeal to norms of Bayesian decision theory to rule out gerrymandered semantic interpretations. Instead, we must rely on *substantive* norms of practical and epistemic rationality. These substantive rationality norms will impose different empirical standards of eliteness for specific predicates, depending on whether their core conceptual role is involved in practical or theoretical reasoning. In the case of thin normative predicates like 'is N', Williams suggests,

[12] For this worry, see Schroeter and Schroeter (2013). A somewhat different worry is that van Roojen's notion of discipline-relative eliteness illicitly relies on an *intentional understanding* of what particular disciplines are and which sorts of property they represent, which would undercut the explanatory value of the account (Williams 2018).

[13] In taking this approach to reference magnetism, Williams is building on work by Schwarz (2014), Weatherson (2013).

the normative role can by itself suffice to fix reference to a specific empirically instantiated property.

To understand the motivations for this approach, it's helpful to start with norms governing epistemic rationality. What makes it the case that your predicate 'is green' attributes the property of *being green* rather than the gerrymandered property of *being grue*? As Nelson Goodman (1995) pointed out, purely formal norms of epistemic rationality, like Bayesianism, treat these two properties as equally good bases for empirical induction. So a Bayesian version of radical interpretation will treat these properties as equally elite referential candidates—resulting in radical indeterminacy of reference for your predicate 'is green'. However, the proper response to Goodman's new puzzle of induction is to posit further, substantive norms of substantive rationality, which require empirical induction to rely on *objectively projectable properties*—that is, properties which support counterfactuals, underwrite dispositions and other varieties of scientific necessity, are confirmed by their instances, and so on. Once this substantive epistemic norm is incorporated into the norms governing radical interpretation, the interpretation function can discriminate against gerrymandered properties like *being grue* (Schwarz 2014; Weatherson 2013). In particular, if the core conceptual role of your predicate 'is green' includes its epistemic role in induction, ER_I, then substantive radical interpretation will strongly favor assigning *being green* over *being grue* as its reference. But for terms *not* governed by any such inductive role—like 'is grue', which has a stipulative theoretical definition—substantive epistemic norms do not favor assigning a projectable property as reference. On Williams' approach, then, reference magnetism is grounded in substantive norms of epistemic rationality, which are relevant to interpretation of particular predicates in virtue of the core conceptual roles governing those predicates.

Williams then extends this model to thin normative terms, by positing substantive norms of practical rationality. Just as it's objectively correct to rely on projectable properties in induction, it is objectively correct to rely on specific normative standards in deciding what to do—for example, substantive norms of practical rationality might mandate:

(K) *Everyone should act according to the categorical imperative.*

Suppose that the specific conventional role (CR) governing your thin normative term 'is morally wrong' involves the following rules (cf. Williams 2018: 43):

(CR-wrong):
(CR-a) Whenever you judge 'a is wrong', you should *blame* a, and
(CR-b) Whenever you judge 'a is not wrong' you should *not blame* a.

Substantive radical interpretation then seeks to construe this CR-wrong as objectively rational (or 'reason responsive') by assigning the property of *violating the categorical imperative* as the reference of your term (2018: §2.3).

Of course, in addition to CR-wrong you also have various *epistemic dispositions* to classify actions as falling into the extension of your predicate 'is wrong'. These dispositions include SCR, the coherent inductive and theoretical practices involved in treating a predicate as having a stable classificatory role. But SCR also entails that you will have *topic-specific* inductive and theoretical practices that constrain your reasoning about which actions fall into the extension of 'is wrong' in specific circumstances.[14] You cannot responsibly update and refine your beliefs about which actions are wrong unless you have some relatively stable topic-specific commitments about how to figure out which things have that property: for example, paradigm cases, theoretical commitments, coherence constraints, methodological presumptions, etc. that are distinctive of this topic. In short, SCR entails that you will have something like the disciplinary standards posited by van Roojen, ER_{m}.

Williams's defense of Referential Normativity depends on the interpretation function treating these epistemic roles as strictly irrelevant to the assignment of semantic contents to normative predicates. Otherwise, individuals whose predicates share the same normative role, like Anne, Beth, Chris and Doug, may associate that predicate with distinct epistemic roles. And the property that best rationalizes the normative role may not match the property that rationalizes the epistemic role. For instance, Doug's use of the term 'is wrong' may conform to CR-wrong, which supports the assignment of *violating the categorical imperative* as its reference; but his inductive and theoretical practices ER_{D} support the assignment of *maximizing pizza consumption*. In such cases, Williams argues, substantive rationalizing interpretation should attribute referential indeterminacy (2018: §4.1).

So securing Referential Normativity requires that we lean very heavily on the notion that CR-wrong plays the psychological role of a *stipulative definition*: CR-wrong is a 'basic disposition' treated as analytic (cf. 2018: 62). In contrast, ER_{M} plays the role of a contingent synthetic commitment: it's a 'derived disposition', and as such plays no role in semantic interpretation (2018: 63). Clearly, Williams' account rests on very strong psychological assumptions about the nature of linguistic competence.

We cannot fully evaluate the merits of Williams's account here. It raises some important questions, about the psychological basis of linguistic competence,

[14] The relevant epistemic practices may be determined by stipulation, by your current mental dispositions, or they may be fixed in part by your past history, linguistic community, or your environment.

about the metaphysical grounding the objective norms of practical rationality, and about our epistemic access to those norms. But we'd like to highlight one general worry about the shape of the account.

The basic idea of Referential Normativity is that semantic rules *that merely require you to act or feel in certain ways when you apply a predicate* will suffice for the interpretation function to assign a semantic content that:

(i) attributes stable referential function to that predicate, and
(ii) singles out a specific empirically instantiated property as its reference.

Williams claims that an independently plausible account of semantic interpretation—*substantive radical interpretation*—will secure (ii). According to Williams, a predicate stipulatively defined by a single semantic rule, CR-wrong, *must* be interpreted as fixing reference to a specific property. Any other interpretation would be unacceptable, since it would fail to construe an individual's use of this predicate as practically rational.

Rather than directly contesting Williams' claim about reference-fixing, we'd like to focus instead on the prior question concerning stable referential purport. In previous sections, we argued that a conventional normative role (CNR)—which requires judgments applying the predicate to be paired with specific patterns of deliberation, motivation, choice and action—will not suffice by itself to secure stable referential purport. The argument turned on two core theoretical roles played by content attributions: *psychological explanation* and *normative assessment*. CNR is consistent with arbitrary variation in competent speakers' epistemic dispositions to apply the predicate to cases. As a consequence, attributing stable referential purport is significantly less useful for the purposes of psychological explanation than attributing a purely expressive function. Moreover, CNR by itself does not seem to favor stable referential purport on normative grounds. A normative role like CR-wrong just tells competent users of a predicate *how to feel* when they attribute that predicate to something. Why shouldn't we take CR-wrong to warrant assigning a purely expressive function to the predicate?

Williams might reply that there are norms of substantive rationality that require us to feel blame towards actions that violate the categorical imperative. But this point is not sufficient to explain why we should think that this predicate has the function of representing a property in the first place. Clearly it would be wrong to claim that *any* semantic rule that invokes the property of *feeling blame* has the semantic function of stably representing that property.

For instance, we could plausibly formulate a semantic rule for a predicate, 'is B', that imbues that predicate with the semantic function of conventionally expressing the speaker's own attitude of blame directed towards any actions it's predicated of. As a model, consider David Kaplan's example of expressives like 'damn', 'ouch' and 'oops': these words are governed by conventional rules that require they be used only when the speaker is in an appropriate mental state (having a derogatory attitude, a sudden sharp pain, an observation of a minor mishap, respectively). But such terms are not plausibly interpreted as contributing to the truth-conditions of the utterances in which they figure (Kaplan 1997). Instead, the use of these words *conventionally implicates* that a speaker is in the relevant mental state (Potts 2005). So we should be able to introduce an expression that has the semantic function of conventionally implicating the speaker's attitude of blame directed at a particular object, without attributing any property to that object. The content we want to express is the expressive/performative content: *x is hereby blamed by me*. Plausibly, we could introduce semantic rules for a predicate, 'is B', that has this semantic function.

We submit that the conventional rules needed to secure this interpretation would look exactly like the rules, (CR-wrong), that Williams proposes for 'is wrong':

(CR-B):
(B-a) Whenever you judge 'a is B', you should *blame* a, and
(B-b) Whenever you judge 'a is not B' you should *not blame* a.

Intuitively, these two conventions seem to suffice for the predicate 'is B' to conventionally implicate that you blame a whenever you assert 'a is B'. So an adequate theory of interpretation should assign the predicate (at least) this expressive semantic function. The key question for our purposes is whether an adequate theory of interpretation must *in addition* attribute stable referential purport. Do these rules suffice to ensure that the predicate 'is B' has the semantic function of attributing the very same property on each occasion of use?

Prima facie, this interpretation is not warranted by CR-B. It seems perfectly possible to introduce an idiom with rules like these that do nothing more than conventionally express the speaker's subjective states. (Compare: Billy says 'Spinach is yuck!' and Dad replies 'No, spinach is *not* yuck!') The case for representational purport could perhaps be strengthened by elaborating the logical role LR 'is B' plays in combinatorial semantics and in inferential patterns. But as we noted in §4, expressivists have developed new resources for explaining these features without resorting to stable representational purport; and relativists, contextualists and localists can also plausibly deny this claim. The lesson we draw is that the rules CR-B + LR will not suffice to secure the stable referential purport of a term—even when we posit substantive norms of practical rationality. More needs to be added to the conventional semantic rules governing the predicate to secure a referential interpretation. This lesson, moreover, generalizes to any version of CNR + LR.

In §4 we suggested that what's needed to secure stable referential purport is a stable classificatory role SCR. SCR involves a suite of epistemic dispositions – classificatory, inductive, and theoretical—characteristic of accumulating a body of information about a specific feature of the world and reflectively monitoring the coherence of the accumulated information. Once a predicate is governed by SCR there is strong interpretive pressure to assign stable referential purport.

However, adding SCR to CR-B + LR will tend to undermine Ardent Normative Realist's claim that a predicate 'is B' stably picks out an empirically instantiated property. As we've seen, Williams himself thinks that adding the sorts of epistemic dispositions involved in SCR to the basic semantic rules governing a normative predicate will lead to semantic indeterminacy. So it seems that securing stable referential purport will prevent determinate reference-fixing on Williams's account.

A further worry that arises for Williams is whether his account can succeed in making *normatively relevant* properties like *violating the categorical imperative* highly elite referential candidates, merely on the basis of a semantic rule like CR-B. According to Williams, substantive radical interpretation requires us to construe the individual's reliance on CR-B as practically rational; and the substantive norms of practical rationality set objectively valid standards for which actions ought to be blamed—or at least they make some standards more elite than others (e.g., *violating the categorical imperative, failing to maximize utility, expressing vicious dispositions,* etc.). So the fact that an individual has a predicate governed in part by CR-B makes such properties elite referential candidates for the predicate 'is B'.

But it's not obvious why these normatively relevant properties should be privileged as referential candidates over an egocentric, purely descriptive property, like *being among the things I blame*. If we were to try to introduce semantic rules that would secure reference to this property, we could do so by introducing (in addition to other rules like LR and SCR) rules like CR-B. But if Williams is right that radical interpretation of CR-B makes normatively relevant properties highly elite referential candidates, it seems that our efforts to introduce a predicate to pick out an egocentric descriptive predicate are bound to fail: the property of *violating the categorical imperative* will be highly magnetic in virtue of CR-B. But intuitively, there seems nothing in CR-B that *should* privilege a normatively relevant property over an egocentric descriptive property as the reference: the rules simply tell you that the attitude of blame should co-vary with the applicability of the predicate. In particular, these rules *don't* require you to check whether blame is normatively *warranted* or *deserved*. So why should the topic-neutral norms of rationalizing interpretation impose this requirement? The proposed interpretive principle favoring normative reference magnets seems to be implausibly skewed towards attributing normatively relevant properties, even when the semantic rules governing an expression have nothing to do with reasoning about what's practically rational.[15]

In this section, we've considered whether an appeal to reference magnetism might vindicate Referential Normativity, by imposing extra empirical constraints on the relative eliteness of referential candidates. We argued that most versions of normative reference magnets are not suitable for this role, since relative eliteness only figures in securing reference after substantive epistemic norms have narrowed down the field of eligible referential candidates. This is true both of the two-stage theories floated by Edwards, and Dunaway and McPherson, and of van Roojen's discipline-specific approach to reference determination. In contrast, Williams seeks to secure stable reference purely on the basis of CR-wrong. There are many aspects of this account that are still to be developed. However, we've argued that no plausible interpretation function can assign stable referential purport solely on the basis of CR-wrong. In addition, we've suggested that Williams's account of normative reference magnetism imposes an implausible interpretive bias towards normatively relevant properties that will result in implausible interpretations for seemingly non-normative expressions.

7. Conclusion and the Further Question

Referential Normativity is attractive to Normative Realists of all stripes because it promises to vindicate our intuitions about the scope for substantive epistemic disagreement over the empirical instantiation conditions of normative properties. As long as individuals associate 'is N' with a specific normative role, Referential Normativity says they are guaranteed to attribute the very same property—irrespective of how radically they may disagree about the nature of that property or its empirical instantiation conditions. Moreover, as Eklund points out, Ardent Normative Realists should find the thesis particularly attractive, since it explains how we can raise substantive questions and enter into genuine disagreements about the nature of mind-independent normative reality.

[15] We have argued that Ralph Wedgwood's (2001) conceptual role semantics for normative terms suffers from a similar problem (Schroeter & Schroeter 2003).

However, we have argued that plausible general principles of semantic interpretation cannot secure stable reference solely on the basis of the action-guiding normative role of normative predicates. The practical aspects of a predicate's conceptual role won't even suffice for *stable referential purport*—much less single out a specific empirically instantiated property as the reference. To secure referential purport and fix reference, semantic interpretation must take into account specific *epistemic roles* played by a normative term in the mental economies of an individual (and their community). Our overall conclusion, then, is that Referential Normativity is a false hope.

Our critique of Referential Normativity was based on three constraints on an adequate theory of semantic interpretation. First, the semantic content of expressions in an individual's idiolect (and the associated conceptual contents) must be useful for the purposes of *psychological explanation*. We must be able to use our semantic assignments in predicting and explaining individuals' reasoning and action. Second, the assigned contents must set plausible standards of *normative assessment*. In particular, the truth-conditions of belief must not set arbitrary epistemic standards: a plausible assignment of truth-conditions must be justifiable on the basis of the individuals' own conceptual practices. Third, the contents assigned must be justifiable on the basis of *general principles of interpretation* that apply to all expressions in all idiolects. A topic-specific theory of interpretation geared toward normative predicates is inherently incomplete, and it can always be gerrymandered to achieve the desired results. We've argued that various efforts to vindicate Referential Normativity conflict with one or more of these constraints on plausible interpretation.

How worrying is this conclusion for Ardent Normative Realists? In this paper, we wish to remain neutral on this question. Eklund suggests that without Referential Normativity, the Ardent Realist cannot explain how we are able to raise the crucial question of *which property captures the true normative joints of reality*. Without Referential Normativity this Further Question about the true nature of normative reality cannot be expressed—it's ineffable. But is Eklund right that the failure of Referential Normativity commits Ardent Realists to the ineffability of the Further Question? And if he is right on this point, how problematic would this ineffability be for Ardent Realists? *These* further questions we'll leave for another occasion.

Acknowledgments

Thanks to Matti Eklund, Tristram McPherson, Bill FitzPatrick and three anonymous referees for helpful comments on previous drafts.

References

Baker, Derek and Jack Woods (2015). How Expressivists Can and Should Explain Inconsistency. *Ethics*, *125*(2), 391–424.

Blackburn, Simon (1998). *Ruling Passions*. Oxford University Press.

Brandom, Robert (1994). *Making It Explicit: Reasoning, Representing and Discursive Commitment*. Harvard University Press.

Burge, Tyler (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*(1), 73–121.

Burge, Tyler (1986). Individualism and Psychology. *Philosophical Review*, *95*(1), 3–45.

Burge, Tyler (1986). Intellectual Norms and Foundations of Mind. *Journal of Philosophy*, *83*(12), 697–720.

Chrisman, Matthew (2012). On the Meaning of 'Ought'. In Russ Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 7, 304–32). Oxford University Press.

Chrisman, Matthew (2015). *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*. Oxford University Press.

Davidson, Donald (1973). Radical Interpretation. *Dialectica*, *27*(3/4), 313–28.

Davidson, Donald (1987). Knowing One's Own Mind. *Proceedings of the American Philosophical Association*, *60*(3), 441–58.

Devitt, Michael (1981). *Designation*. Columbia University Press.

Dreier, James (1990). Internalism and Speaker Relativism. *Ethics*, *101*(1), 6–26.

Dretske, Fred (1988). *Explaining Behavior*. MIT Press.

Dunaway, Billy and Tristram McPherson (2016). Reference Magnetism as a Solution to the Moral Twin Earth Problem. *Ergo*, *3*(25), 639–79.

Edwards, Douglas (2013). The Eligibility of Ethical Naturalism. *Pacific Philosophical Quarterly*, *94*(1), 1–18.

Eklund, Matti (2017). *Choosing Normative Concepts*. Oxford University Press.

Enoch, David (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.

Finlay, Stephen (2014). *Confusion of Tongues: A Theory of Normative Language*. Oxford University Press.

FitzPatrick, William (2008). Robust Ethical Realism, Non-Naturalism and Normativity. In Russ Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 3, 159–205). Oxford University Press.

FitzPatrick, William (2018). Representing Ethical Reality: A Guide for Wordly Non-Naturalists. *Canadian Journal of Philosophy*, *48*(3–4), 548–68.

Frege, Gottlob (1997). On *Sinn* and *Bedeutung*. In Michael Beany (Ed.), *The Frege Reader* (151–71). Blackwell. (Original work published 1892)

Gert, Joshua (2018). Neo-Pragmatism, Morality, and the Specification Problem. *Canadian Journal of Philosophy*, *97*(3–4), 447–67.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Harvard University Press.

Gibbard, Allan (2003). *Thinking How to Live*. Harvard University Press.

Goodman, Nelson (1995). The New Riddle of Induction. In *Fact, Fiction and Forecast*. Harvard University Press.

Harman, Gilbert (1975). Moral Relativism Defended. *Philosophical Review*, *84*(1), 3–22.

Joyce, Richard (2001). *The Myth of Morality*. Cambridge University Press.

Kalderon, Mark Eli (2005). *Moral Fictionalism*. Oxford University Press.

Kaplan, David (1997). The Meaning of 'Ouch' and 'Oops'. In *Howison Lecture in Philosophy delivered at UC Berkeley*. http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf: transcribed by E. Coppock.

Kölbel, Max (2002). *Truth without Objectivity*. Routledge.

Lewis, David (1974). Radical Interpretation. *Synthèse*, *27*(3/4), 331–44.

Lewis, David (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, *61*(4), 343–77.

Lewis, David (1984). Putnam's Paradox. *Australasian Journal of Philosophy*, *62*(3), 221–36.

MacFarlane, John (2014). *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford University Press.

Mackie, John (1977). *Ethics: Inventing Right and Wrong*. Penguin.

McPherson, Tristram (2019). Supervenience in Ethics. In Edward N. Zalta (Ed.), The *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/win2019/entries/supervenience-ethics/

Millikan, Ruth Garrett (1984). *Language, Thought, and Other Biological Categories*. MIT Press.

Millikan, Ruth Garrett (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.

Pérez Carballo, Alejandro (2015). Semantic Hermeneutics. In Alexis Burgess and Brett Sherman (Eds.), *Metasemantics: New Essays on the Foundations of Meaning* (119–46). Oxford University Press.

Potts, Christopher (2005). *The Logic of Conventional Implicature*. Oxford University Press.

Price, Huw (2011). *Naturalism without Mirrors*. Oxford University Press.

Putnam, Hilary (1970). Is Semantics Possible? In H. E. Kiefer and M. K. Munitz (Eds.), *Language, Belief and Metaphysics* (50–63). SUNY Press.

Putnam, Hilary (1975). The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science*, *7*, 131–93.

Putnam, Hilary (1980). Models and Reality. *Journal of Symbolic Logic*, *45*(3), 464–82.

Putnam, Hilary (1981). *Reason, Truth and History*. Cambridge University Press.

Rayo, Agustín (2013). A Plea for Semantic Localism. *Noûs*, *47*(4), 647–79.

Schroeter, Laura and François Schroeter (2003). A Slim Semantics for Thin Moral Terms? *Australasian Journal of Philosophy*, *81*(2), 191–207.

Schroeter, Laura and François Schroeter (2013). Normative Realism: Co-Reference without Convergence? *Philosophers' Imprint*, *13*(13), 1–24.

Schroeter, Laura and François Schroeter (2017). Metasemantics and Metaethics. In Tristam McPherson and David Plunkett (Eds.), *The Routledge Companion to Metaethics* (519–35). Routledge.

Schroeter, Laura and François Schroeter (2018). Keeping Track of What's Right. *Canadian Journal of Philosophy*, *48*(3–4), 489–509.

Schroeter, Laura and François Schroeter (2019). The Generalized Integration Challenge in Metaethics. *Noûs*, *53*(1), 192–223.

Schwarz, Wolfgang (2014). Against Magnetism. *Australasian Journal of Philosophy*, *92*(1), 17–36.

Streumer, Bart (2017). *Unbelievable Errors: An Error Theory about All Normative Judgments*. Oxford University Press.

Sundell, Timothy (2012). Disagreement, Error, and an Alternative to Reference Magnetism. *Australasian Journal of Philosophy*, *90*(4), 743–59.

van Roojen, Mark (2006). Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument. In Russ Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 1, 161–94). Oxford University Press

van Roojen, Mark (2018). Moral Cognitivism vs. Non-Cognitivism. In Edward N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/fall2018/entries/moral-cognitivism/

Weatherson, Brian (2013). The Role of Naturalness in Lewis's Theory of Meaning. *Journal for the History of Analytic Philosophy*, *1*(10), 1–19.

Wedgwood, Ralph (2001). Conceptual Role Semantics for Moral Terms. *Philosophical Review*, *110*(1), 1–30.

Williams, J. R. G. (2018). Normative Reference Magnets. *Philosophical Review*, *127*(1), 41–71.

Williamson, Timothy (2007). *The Philosophy of Philosophy*. Blackwell.