

The Limits of Normative Detachment¹

Nishi Shah

Amherst College

Consider another picture of what it would be for a demand to be ‘objectively valid’. It is Kant’s own picture. According to this, a demand will be inescapable in the required sense if it is one that a rational agent must accept if he is to be a rational agent. It is, to use one of Kant’s favourite metaphors, *self-addressed* by any rational agent. Kant was wrong, in my view, in supposing that the fundamental demands of morality were objective in this sense, but that is not the immediate point, which is that the conception deploys an intelligible and adequate sense of objectivity. It seems to have little to do with those demands being part of the fabric of the world; or, at any rate, they will be no more or less so than the demands of logic – which was, of course, part of Kant’s point.

-- Bernard Williams, “Ethics and the fabric of the world”²

John Mackie famously argued that the objective purport of moral discourse requires that values be part of the fabric of the universe, but that the universe, at least as it has been disclosed to us by the natural sciences, contains no such “queer” properties. Nor have the natural sciences disclosed any perceptual or other capacities that would allow us to detect the presence of such properties, even if they did exist. Mackie thus claimed that moral judgments are not what they sometimes appear to be: warranted, true statements of objective fact.³ Part of the interest of the Kantian strategy that Williams describes is that it holds out the promise of vindicating the objectivity of moral discourse without requiring the existence of any such queer entities.⁴ Moral statements attain objectivity, not by accurately reflecting aspects of normative reality, but by expressing inescapable commitments of practical agency.⁵

¹ Acknowledgements.

² Williams (1995), 172 – 181.

³ Mackie (1977).

⁴ See Velleman (2009), 115-17, for a similar interpretation of Williams’s quotation.

⁵ Williams puts the Kantian claim in terms of the commitments of rational agency, not practical agency. But the term ‘rationality’ is potentially ambiguous. According to one usage, rationality is defined in terms of normative truth. This cannot be the conception of rationality that Williams is attributing to the Kantian because demonstrating that rationality in this sense requires the acceptance of some normative claim need not differ from demonstrating that the normative claim accurately reflects an aspect of normative reality.

Elsewhere I have attempted to apply the Kantian strategy to vindicate the objectivity of the fundamental doxastic norm of truth. The question that expresses the aim of doxastic deliberation, *whether to believe that p*, immediately gives way to the question *whether p is true*. I have argued that the best explanation of this feature of doxastic deliberation is that in deliberating whether to *believe that p* one must, in virtue of exercising the concept of belief, apply the norm of truth to one's reasoning.⁶ More specifically, one must apply the norm:

Believing that *p* is correct if and only if *p*.⁷

By showing that doxastic deliberation, a central activity of epistemic agency, involves a commitment to this norm, this explanation reveals that the norm, and whatever norms can be derived from it, are objectively valid in the sense Williams attributes to Kant: one's application of them doesn't depend on any contingent subjective features of one's psychology not shared with all epistemic agents. In order to engage in those activities essential to the exercise of one's epistemic agency, one must apply this norm. I will refer to this kind of objectivity as Kantian objectivity and the class of arguments such as mine that aim to establish Kantian objectivity the Kantian strategy.⁸ My use of the label 'Kantian' is meant to place these arguments in a certain tradition inspired by reading Kant, not to attribute these arguments to Kant himself.

The Kantian's conception of rationality is a procedural one defined in terms independent of normative truth. To avoid this confusion about the meaning of 'rationality', I have avoided using the term in this paper.

⁶ See Shah (2003) and Shah and Velleman (2005).

⁷ The claim that believing that *p* is correct iff *p* is not equivalent to the claim that believing that *p* is true iff *p*. Strictly speaking, propositions, not beliefs, are true or false; beliefs are psychological states whose contents are true or false but are themselves neither true nor false. My claim is that it is a conceptual truth that beliefs are normatively assessable as correct or incorrect in virtue of the non-normative property that their propositional contents have of being true or false.

⁸ Christine Korsgaard employs this strategy in *The Sources of Normativity* (1996) and "The Normativity of Instrumental Reason," (1997). David Velleman also employs a version of this strategy in "The Possibility of Practical Reason," (2000) although he claims only that the relevant norm must be embodied in a sub-agential aim that guides the relevant activity, not that the agent must himself accept the relevant norm in order to engage in the activity. This version of the Kantian strategy raises different questions than the ones I pursue here. See Shah (2003) for a critical discussion of Velleman's account as applied to doxastic normativity.

All versions of the Kantian strategy attempt to show that there are certain normative commitments that are constitutive of being an agent. The problem is that, so described, the Kantian strategy does not seem to preclude the possibility that being an agent requires one to hold mistaken beliefs. For all that the Kantian strategy says, the commitments that are constitutive of agency are beliefs, and nothing about their inescapability seems to rule out the possibility that they are uniformly false. Anyone familiar with the literature on free will certainly will have read pronouncements of such inescapable error. Error-theorists about free will claim that exercising our agency requires making a mistake about our own abilities: in order to engage in practical deliberation one must believe that one's choices are free in a very strong sense, but this type of freedom, they argue, is incompatible with the laws of nature. Mackie himself appears to have thought that exercising human agency involves a different kind of mistake. Even though he claimed that moral judgments express beliefs about an illusory realm of non-natural facts, he did not believe that it was an option for us to renounce moral beliefs altogether. Mackie's argument for an error-theory of moral judgments thus appears untouched by the Kantian position about objectivity that Williams describes.⁹ That position just does not speak to the semantic, metaphysical, and epistemological questions pursued by traditional metaethicists such as Mackie, nor does it show that there is anything defective about those questions.¹⁰

The Kantian strategy thus appears to be metaethically neutral. Whether this appearance of neutrality is veridical depends on whether the Kantian strategy really does

⁹ Mackie does say that his denial of objective values is the denial of the objective validity of a categorically imperatival element in ethical judgments. It is unclear what Mackie had in mind here. I don't see that he could have meant to deny that ethical norms are objectively valid in the Kantian sense that Williams describes. Denying Kantian objectivity would require showing, not that our normative judgments are systematically false, as Mackie argued, but that it is possible for us to shed these norms (while retaining our agency). But Mackie couldn't have understood his argument to have shown this, since he thought that it wasn't an option for us to give up ethical judgments. This does leave a puzzle as to what Mackie meant to reject by his denial that ethical demands are objectively valid.

¹⁰ Williams claims that the Kantian offers an anti-realist objective grounding of morality. But as far as I can see a realist is free to accept that moral judgments have Kantian objectivity. Realists claim that moral judgments are objective in the sense that their truth doesn't depend on what anyone believes about them. They thus claim that ethical judgments are capable of objective *truth*. Williams's Kantian argues that moral judgments are objective in the sense that they are inescapable for rational agents. There is no reason that a moral judgment couldn't both be objectively true and inescapable. The realist and Kantian, at least at this level, are not advancing rival views about the same attribute of moral judgments, but are advancing views about two different but compatible attributes of moral judgments.

leave open the possibility of an error-theory. Nonetheless, once it has been shown that ethics has a foundation in norms that are inescapable for any practical agent, it may appear that such a ‘metaethical’ view loses its point. After all, if there are norms that we as agents must accept, what would motivate us to ask whether we accept them in the form of beliefs, or whether such beliefs have any chance of being true? Certainly it cannot be a desire to find an authoritative basis from which to settle normative disputes. If the Kantian argument works, it shows us that there are norms that must be compelling to any and all agents. This means that these norms have an authority that cannot be disputed, no matter whether utterances of them are truth-apt or possibly true. These norms thus function as objective standards in adjudicating normative disagreements.

But the concern that might lead us to ask further metaethical questions is not the threat of relativism or nihilism—that there are no universally valid normative standards¹¹—but of being duped. If we are worried by Mackie’s argument, it is because we are worried that our ethical practices are based on a mistake. In the same way that witchcraft is a practice based on false attributions of supernatural powers to people, if Mackie is right, praising and blaming people is a practice based on false attributions of objectively prescriptive properties to their actions. Being told that everyone is mistaken, or even that it is impossible for anyone to fail to be mistaken, will not assuage this worry. In fact, it may deepen it. That we have no choice but to acquiesce in a set of judgments would give us all the more reason to hope that the judgments are true, since if they are not, there is nothing we can do about it—we will be compelled to believe in a fantasy.¹²

¹¹ The standards are valid, not merely shared, because they are inescapable for agents. Their validity therefore consists in the fact that they are incapable of being invalidated by agents, being beyond questioning by them.

¹² My concern about the status of the Kantian Strategy is different than, though consistent with, that expressed by David Enoch (2006). He argues that the Kantian Strategy fails to validate the norms it claims are constitutive of agency because it fails to establish that we ought to be agents. Enoch thinks that the most that the Kantian can establish are conditionals of the following form:

If one ought to be an agent, and being an agent requires accepting that x is f, where f is a normative predicate, then x is f.

My concern is whether the Kantian can even establish this conditional. To see the difference between our concerns, suppose that there are sufficiently weighty pragmatic benefits to be had from being an agent to make it the case that one ought to be an agent. Suppose further that, for example, Korsgaard (1996) is correct that to be an agent one must value humanity. From these claims it does not follow that humanity

I want to investigate whether there is an interpretation of the Kantian strategy for securing normative objectivity that can answer or at least mollify the concern that our inescapable acceptance of certain norms is a collective mistake. Is there a way of understanding the Kantian strategy as a method of establishing the truth of certain normative judgments? If not, might the Kantian strategy be used in some other way to undermine the threat that we are duped in our judgments about such fundamental norms? If the Kantian strategy successfully either provided a method of establishing normative truths, or demonstrated that a demand for such a demonstration is misconceived, would there remain any lingering skeptical doubts concerning normative discourse?

The rest of this paper divides into two sections. The first section explores a tempting line of thought that leads to a constructivist interpretation of the Kantian strategy. Constructivism, if true, would justify treating the Kantian strategy as a method for arriving at normative truths. I will argue, though, that a constructivist interpretation is of dubious coherence, and in any case is unavailable to those who seek to apply the Kantian strategy across the board to norms of belief and action alike. The second section examines a more defensive strategy that the Kantian might employ against an error-theorist. I describe an argument that attempts to show that, even absent a demonstration that it is a method of establishing normative truths, the Kantian strategy, if applied to both norms of belief and action, is invulnerable to any completely general argument that all of our normative judgments are false. I conclude with some reflections on the implications of this argument for the practice of metaethics.

I

really is valuable. Similarly, even if Pascal's Wager provides reasons to be religious, and being religious requires believing that God exists, it does not follow that God exists or even that there is any evidence that God exists, as Pascal himself fully appreciated.

Before moving forward, let me summarize the discussion to this point. I started by introducing the Kantian strategy of securing a kind of normative objectivity by showing that a commitment to certain norms follows from the fact that one is a practical or doxastic agent. I then raised questions about the metaethical status of these arguments. I claimed that while they might demonstrate that there are certain norms that cannot be questioned in normative discussion and are therefore in a sense valid for all agents, they do not respond to the kind of anxiety that is raised by Mackie's arguments for an error-theory of normative discourse. One form this anxiety takes is the worry that we are under a collective, if inescapable, illusion: our normative judgments, like judgments that describe certain people as witches, predicate non-existent properties to objects and events¹³ and thus express beliefs that are systematically false.

How should the Kantian respond to this threat? Is there an interpretation of the Kantian strategy that can relieve this anxiety by showing that it is a method for establishing the truth of certain normative judgments?

The Kantian strategy, if it succeeds, accomplishes one very important task, which is to show that our fundamental normative judgments are deliberately invulnerable. To show that a certain normative judgment is inescapable for agents implies that from an agent's deliberative perspective, it could never be an open question whether to accept the judgment. One could never be in the position of both asking the question whether to accept the judgment, a question that one can only ask if one is an agent, while withholding assent from the judgment.

Might this result imply that our fundamental normative judgments are *epistemically* justified? An initial problem is this: even assuming that showing that a judgment has this inescapable status is a way of justifying it in some sense, this form of justification does nothing to show that the judgments that are thus justified are expressive of beliefs rather than some kind of non-cognitive state, nor that these judgments are true or likely to be

¹³ The categories of object and event, as I am using them, include persons and actions.

true. This form of justification thus does not seem to shed any light on the *epistemic* credentials of the judgments that are so justified.

If normative judgments were expressions of non-cognitive states there would be no substantive question whether these judgments have or lack epistemic credentials. Because normative judgments would not express cognitive states with robust truth-conditional contents, there would be no sense to be made of justifying these judgments by adducing evidence that they are true.¹⁴ This response to the error-theorist would not be, however, a way of showing that the Kantian strategy itself undermines the error-theorist's position, since it in no way depends on the success of the Kantian strategy. If the non-cognitivist is correct, an error-theory is false not because it implies that there are no set of inescapable normative judgments, but because it mistakenly interprets normative judgments as beliefs. My concern in this paper is whether the Kantian strategy can itself be used to block the error-theorist's conclusion that our normative judgments are systematically false, so I will not explore the possibility of a Kantian non-cognitivist position here.¹⁵

The Kantian I now have in mind thinks that non-cognitivism has abandoned the main aspiration of metaethics. This aspiration is not merely to give some coherent interpretation of our normative practice, but to show how our normative judgments could be what they sometimes seem to be, true and epistemically justified beliefs about a distinctive subject matter. While Kantians think that non-cognitivists have ducked this task by interpreting normative judgments as having no distinctive subject matter, they think that non-reductive realists, by interpreting normative judgments to be about an ethereal realm of independent normative facts, have made the task impossible: there is no

¹⁴ Even modern expressivist heirs to noncognitivism such as Gibbard (2003) and Blackburn (1998) are committed to saying that it is inapt to talk of the epistemic justification of the states expressed by normative judgments when these judgments are characterized solely in terms of their most explanatorily fundamental description. Of course, if they are right, these attitudes, in virtue of being embedded in a practice that exhibits the requisite logical discipline, will correctly be classified as beliefs about whose truth or falsity we can dispute. But expressivists are committed to claiming that there is an alternative true characterization of these states that makes no mention of their truth-conditions, and that this non-cognitive characterization is the one adverted to in the best explanation of the fundamental features of moral practice. These features of course include the very fact that moral practice exhibits the logical discipline that makes it appropriate to apply the truth predicate to moral judgments and thus to speak of them as expressing beliefs.

¹⁵ I will briefly discuss how non-cognitivism might be combined with the Kantian strategy at the end of the paper.

way that agents like us, with our limited empirical means, could gain epistemic access to normative facts so understood.¹⁶ Certainly showing that a certain normative judgment is inescapable for rational agents is not a way of demonstrating that it is true of an independent normative realm, but if non-reductive realism were true, this is exactly what such an argument would have to demonstrate to epistemically justify the judgment.

Our question thus is whether there is an interpretation of normative truth and normative judgment according to which demonstrating that a normative judgment is inescapable for agents is a way of showing that the judgment is a true belief about a distinctive subject matter. Suppose then that normative judgments do not express non-cognitive attitudes, but express beliefs that are in part individuated by their truth-conditional contents. As expressions of beliefs, they are subject to a standard of correctness: they are correct if and only if their contents are true. We thus can ask whether these judgments are justified in the sense determined by their correctness conditions; whether, that is, there is good reason to think they are true.

Is there a way of interpreting the Kantian strategy as demonstrating that a certain class of normative judgments are true—those that are inescapable for agents—and thus as providing a bona fide epistemic justification of them? This depends on what we understand these judgments to be about. If normative judgments are about facts that are metaphysically independent of the normative judgments of agents in general—norm-judgmental facts, then it is hard to see how demonstrating that agents must make one of these judgments would bear on whether the judgment is true. Such a demonstration would prove something about the nature of agents, not something about a domain of facts that exist independently of agency. If there were a cogent argument that it is a necessary truth about human adults that they accept certain judgments about middle-sized physical objects, this by itself would go no way towards showing that these judgments are true—that middle sized physical objects actually exist.

¹⁶ Or at least non-reductive realists have failed to help us understand how we could have normative knowledge, since all they have given us are uncashed perceptual metaphors such as that we just “see what is right.” At best they have described what needs to be explained rather than explaining it. At worst they have mischaracterized the subject matter of normative judgments in such a way that makes an explanation of it impossible.

However, if normative judgments are about facts that are *constructed* from privileged set of normative judgmental facts, the Kantian strategy may be able to provide us with a method for establishing normative truths. For example, if the truth of our actual normative judgments is a matter of their membership in the set of normative judgments that we would accept in light of full empirical information, then the Kantian strategy, by showing that a certain class of normative judgments is inescapable for us, and thus must be part of any set of normative judgments that we could accept, would epistemically warrant those judgments.¹⁷ If normative facts are constituted by facts about the set of judgments we would accept in light of full empirical information, then showing that a certain class of normative judgments is part of any set of normative judgments that an agent could accept is a way of showing that that class of normative judgments is true.¹⁸

Whether or not the Kantian strategy provides an epistemology for the normative judgments that it establishes as objectively valid thus seems to depend on whether we have *independent* grounds for believing that the relevant normative facts are themselves *constructed* out of facts about the normative judgments we make—norm-judgmental facts. In the end it would appear that we still need an independent metaethical theory to deliver the metaphysics that the epistemological interpretation of the Kantian strategy requires.

¹⁷ ‘Full information’ accounts of the truth-makers of various types of normative judgments (e.g. personal good) have been very popular. Importantly, as we shall see later, most of these accounts specify the truth-makers in terms of the desires, rather than the normative judgments, that we would arrive at in light of full information. See Smith (1994) for an attempt to apply a full-information analysis of this kind to all practical reasons. As far as I know no one has applied this style of analysis in full generality to encompass all normative judgments, practical and doxastic alike.

¹⁸ Recently, Korsgaard (2003) has claimed that normative judgments are not about judgment-independent or judgment-dependent facts, because they are not descriptions of facts at all. But denying that normative judgments describe facts is just to deny that normative judgments express beliefs. This is the noncognitivist’s position. But Korsgaard explicitly rejects noncognitivism. Is there another possibility? I doubt it, but an adequate discussion of Korsgaard’s position would take us too far afield. For an attempt to interpret Korsgaard’s metaethical position see Hussain and Shah (2006).

Is there a way though of understanding the Kantian strategy as itself an argument for constructivism¹⁹? It is important at this point to have a clear understanding of the dialectical situation. We are wondering how showing that certain normative judgments are inescapable would be a way of demonstrating that they are true. If normative facts were metaphysically independent of norm-judgmental facts, as the error-theorist supposes, the Kantian strategy would not provide such a demonstration. Does this mean that to defeat the error-theorist the Kantian needs an independent argument for the claim that normative facts are not metaphysically independent of norm-judgmental facts? This depends on what attitude we are allowed to take towards the Kantian strategy antecedently to engaging in metaphysical inquiry. Certainly there is no reason to accept that our metaphysical commitments in general must be arrived at independently of our epistemological views. If we have a *prima facie* entitlement to treat the Kantian strategy as a procedure for establishing normative truths, then we can use the Kantian strategy to justify claims about the metaphysics of normativity.

While there may be reasons not to grant that the Kantian strategy even has such a *prima facie* epistemic status, I will grant it this status for the sake of argument. By treating the strategy as a procedure for arriving at correct answers to normative questions, we commit ourselves to whatever ontology best explains how this procedure could yield correct answers to normative questions.²⁰ We thus don't need an independent reason for thinking that normative facts are constructed out of norm-judgmental facts; the reason for thinking this just is that these are what normative facts would have to be in order for it to make sense to treat the Kantian strategy the way that we do treat it, as a procedure that yields correct answers to normative questions.

¹⁹ I use the term 'constructivism' with some trepidation. Many Kantians who call themselves constructivists will not recognize the position I am about to describe as their own. But, following Rawls (1971), most Kantians have not aspired to give a fully general metaethical account. I hope that from what I have already said that it is clear why an investigation of the position I am calling 'metaethical constructivism' is of interest whether or not any Kantian actually endorses it.

²⁰ Or, to put things in more clearly epistemic terms, by treating the Kantian strategy as a method of gaining normative knowledge, we commit ourselves to whatever ontology best explains how it could yield such knowledge. Given that we are assuming cognitivism here, it is difficult to see why we should abjure applying this description to the upshot of a sound deployment of this strategy.

On this interpretation of the Kantian strategy, its epistemic credentials don't depend on a distinct metaethical argument for constructivism. Constructivism doesn't justify the epistemic credentials of the Kantian strategy, but is itself justified by its role in an explanation of how the Kantian strategy could be a method for arriving at correct normative judgments. In order for constructivism to play this explanatory role, though, it must constitute a determinate view of normative facts. As I will now argue, this is something we have very good reason to doubt.

Constructivism is a thesis of metaphysical priority: normative facts are grounded in or dependent on norm-judgmental facts. For example, according to moral constructivism, lying is wrong, if it is, because some suitable agent under certain conditions would judge that lying is wrong. The specification of the relevant agent and circumstances will vary depending on what type of constructivist view is being offered. Constructivist views can be represented schematically in the following way:

Metaethical Constructivism:

The property of being F, where 'F' stands for a normative predicate, is such that for any x, if x is F, then x is F because S judges that x is F (under conditions C), where it is left open who S refers to, whether it be an individual a community, all agents, etc..²¹

This is a schema for a determinate view only if there is a specification of the judgmental fact out of which the fact that x is F is constructed. Given that the judgment that determines whether some x is F is supposed to express a belief, it is natural to think that its content is to be individuated in terms of its truth-condition. According to the constructivist schema this means that whether x is F is determined by a type of judgment that has that very same propositional content. According to the schema, therefore, whether F has an extension depends on whether judgments that x is F can be assigned

²¹ For a similar definition of metaethical constructivism, see Street (2008).

truth-conditions, but whether these judgments can be assigned truth-conditions depends on whether F has an extension. As we shall see, this circularity is highly problematic.²²

The problem is not that there cannot be truths of the following form:

x is F iff S believes that x is F (under conditions C)

After all, we do believe that there are facts that we form true beliefs about when we are under certain non-trivially specified conditions, and such bi-conditionals allow us to perspicuously express such claims. That such a bi-conditional expresses a truth about a certain domain of facts does not imply that such facts are constructed or constituted out of the relevant belief facts. To say that the relevant facts are constituted out of such belief facts would be to provide a possible explanation of why these bi-conditionals hold.

What is the problem with turning such a bi-conditional into a constitution claim? The claim that one fact A is constructed out of another fact B implies that there is an asymmetric metaphysical dependence between them: A facts metaphysically depend on B facts but B facts do not metaphysically depend on A facts. It might be thought that I am arguing that the existence of beliefs with normative content is asymmetrically dependent on the existence of normative facts, and therefore that normative facts cannot be constituted out of normative belief facts. But this cannot quite be the problem. If an error-theory of normative judgments were correct, and normative beliefs thus were uniformly false, then there would be normative beliefs but no normative facts. And nothing I have said rules out such a possibility.

The problem emerges when we ask what an agent who believes that x is F believes. What property is an agent predicating of x in believing that x is F? The constructivist

²² If the target of analysis were judgment-tokens rather than judgment-types, there would be no problem specifying the truth-conditions of the judgments, since judgment-tokens, unlike judgment-types, can refer to themselves by means of token-reflexives. For example, my intention to go for a walk may involve the self-referential belief that I will go for a walk as a result of *this very belief (token)*, but this self-reference is unproblematic because the content of the belief involves a token-reflexive element ('this very belief') that secures its reference.

answer is that for something to be F is for someone to judge, and in judging believe, that x is F. It is a consequence of constructivism, then, that a specification of the content of the belief that x is F will make reference to the belief that x is F. The belief that x is F is a belief whose content is that x has the property something has in virtue of someone believing that x is F. 'x is F' thus picks out a determinate state of affairs only if there is a determinate mental state referred to by 'the belief that x is F', but whether 'the belief that x is F' picks out is determinate mental state itself depends on whether a determinate content for it can be fixed. Unless there is some independent way to specify the belief or its content, we are left without either.²³ It thus appears that the constructivist has no way to specify the norm-judgmental facts that he believes constitute the metaphysical ground of normative facts. At the very least, he does not have the resources to distinguish between F-facts and G-facts, if both facts are constructed out of judgmental facts. For example, if facts about moral rightness are constructed out of judgmental facts, presumably so too are facts about moral wrongness. But if the contents of the beliefs that constitute these judgments are indeterminate, there will be no fact of the matter about which judgments are rightness judgments and which are wrongness judgments, and thus no fact of the matter about which acts are right and which are wrong.

One way to try to avoid this problem is to interpret normative facts not as constructed from judgments attributing those very properties, but as constructed from judgments attributing some other normative properties. The view would have the following structure:

Metaethical Constructivism*:

The property of being F, where 'F' stands for a normative predicate, is such that for any x, if x is F, then x is F because S judges that x is G (under conditions C).

If constructivism is to be a fully general account of normative truth, the truth-maker of the new normative judgment that x is G must itself be given a constructivist interpretation. This interpretation must itself be given in terms of judgments of the form x

²³ Compare Boghossian and Velleman's objection to a similar account of color facts, (1989), 89-90.

is G, or in terms of some other normative judgments. If the former, then we are back to the problem we started with, but now about how to understand judgments that x is G. If the latter, we are off on a regress, since we will now have to understand the truth-makers of these new judgments in constructivist terms on pain of collapsing into a non-constructivist view.

Might the constructivist accept such a regress? I don't think so. Remember, the constructivist is not out to establish merely that there are true bi-conditionals linking normative facts and norm-judgmental facts. Such bi-conditionals, if true, entail only that, under certain conditions, normative facts co-vary with corresponding norm-judgmental facts; they do not establish that the normative facts are constructed out of those norm-judgmental facts. In order for normative facts to be constituted by norm-judgmental facts, the norm-judgmental facts must be metaphysically prior to the normative facts. But if there is no way to identify any normative judgments without presupposing some normative facts, this metaphysical priority cannot be established.

Let us assume that there is a way for the constructivist to specify determinate truth-makers for the relevant normative judgments.²⁴ Would constructivism now constitute a determinate metaethical position? Remember the schema for the view:

Metaethical Constructivism:

The property of being F, where 'F' stands for a normative predicate, is such that for any x, if x is F, then x is F because S *judges* that x is F (under conditions C).

Even if the content of the judgment out of which the fact that x is F is to be constructed can somehow be fixed, the fact that it is a *judgment* presents an insurmountable problem. The type of judgment invoked in the above constructivist schema expresses a belief. The concept of belief though, according to the Kantian argument I have given elsewhere, is

²⁴ See Velleman (2000b) for an attempt to solve this problem as it arises in the context of specifying the content of joint agreements.

itself a normative concept. To classify a mental state as a belief that *p* is to accept the following normative judgment:

Believing that *p* is *correct* if and only if *p*

Although I believe that there is distinctively Kantian route to this thesis, other philosophers accept it for independent reasons.²⁵ If this thesis is true, for whatever reason, it is difficult to understand what metaethical constructivism could be.

Metaethical constructivism, remember, is supposed to be a fully general account of normative truth, therefore it needs an interpretation of the truth-makers of all normative judgments, whether those judgments express the acceptance of norms of belief or action. Thus, even if the constructivist can fix a content for the judgment that is meant to provide the basis for construction, he will still need to provide an account of the norm invoked in classifying it as a judgment, that is, as a mental act expressive of a belief. If the normative fact of being correct if and only if its content is true is what distinguishes belief from other attitudes, then the constructivist owes us an account of this fact. Otherwise, for all we know, the fact may be such that we have no epistemic access to it. And if it is, so too is anything out of which it is constructed.

Suppose then we try to analyze belief itself along constructivist lines:

Belief Constructivism:

The property of believing that *p* is such that, for any subject R, if R believes that *p*, it is the case that R believes that *p* because S *believes* that R believes that *p* (under conditions C).

Since belief facts are themselves normative facts, in order to complete this account, the constructivist needs to provide an account of the fact that S believes that R believes that

²⁵ See Wedgwood (2002) and Boghossian (2003). As should be evident, the kind of problem I am about to raise for constructivism can also be generated if, as many believe, meaning is a normative concept.

p . The fact that S believes that R believes that p therefore must be constructed out of further belief facts if a fully general constructivism about normativity is correct. The account thus has the following structure:

The fact that S believes that p is constituted out of the fact that T believes that S believes that p , which is itself constituted out of the fact that U believes that T believes that S believes that p , etc...

We again end up with a regress, this time with belief facts being constructed out of further belief facts, which in turn must be constituted out of further belief facts, and so on *ad infinitum*.

The prospects for metaethical constructivism look bleak. The constructivist claims that normative facts are constituted out of norm-judgmental facts. He runs into problems both specifying the contents of those judgments and recognizing the fact that those judgments express beliefs. In attributing a normative belief, one must either presuppose the existence of independent normative facts or be drawn into a vicious regress.

II

Kantians might have hoped that demonstrating that certain normative judgments are inescapable would epistemically justify those judgments. A constructivist metaethic, by establishing a necessary connection between inescapability and truth, would fulfill this hope: it would underwrite the inference from the inescapability of a normative judgment to its truth. As we have seen, though, a fully general constructivist account of normative truth appears unreachable. Establishing that a normative judgment is inescapable thus does not license us to infer that it is true.

In light of the failure of constructivism, should Kantians feel worried that our fundamental normative judgments, inescapable as they may be, might be systematically false? This depends on whether an error-theory can itself avoid presupposing some

normative truths consistently with acknowledging the success of the Kantian strategy across the board. If an error-theory implies that there are no normative judgments one must accept in order to be an agent, then the Kantian can refute the error-theorist by demonstrating the inescapability of some normative judgments. The topic of this paper has not been to examine whether the Kantian strategy succeeds, but to ask whether its success has any implications for metaethics, specifically whether it implies the falsity of an error-theory. If an error-theory implies that there are no normative judgments that one must accept in order to be an agent, demonstrating that there are normative judgments that have this status will likewise imply that an error-theory is false. This is just the result the Kantian was hoping for.

Unfortunately, we have not yet seen any reason to believe that an error-theory does imply the falsity of any Kantian claims. At first glance, it seems that the Kantian is making a claim about which normative claims one must accept to engage in certain activities, and making the further claim that these activities are inescapable for agents, whereas the error-theorist is making a claim about whether those normative claims are true. These views appear to be making logically unrelated claims. But maybe once we look more closely at the conditions that must be satisfied by a fully general error-theory of normative judgments, we will see that there is something problematic or incoherent about accepting both sets of claims. I will conclude by investigating this possibility.²⁶

What must one do to arrive at a fully general error-theory of normative judgments? It is not sufficient that one assert that sentences used to predicate normative properties to objects are all false. To see this, think about something that we are all error-theorists about, witch-discourse. We think that people who believed (or continue to believe) that certain people are witches were mistaken. To arrive at this claim, we must have been able to attribute these beliefs to certain people. Unless we were able to identify practices in which some people held such beliefs, there would be no sense to be made of our claim that anyone has ever been *mistaken* about the existence of witches. At a minimum, to

²⁶ The argument I am about to give bears similarities to Paul Boghossian's argument against error-theories of linguistic and mental content and Barry Stroud's argument against error-theories of colour. See Boghossian (1990) and Stroud (2000, ch.7) respectively.

arrive at an error-theory about normative judgments one therefore must do the following two things:

- 1) Attribute normative beliefs.
- 2) Demonstrate that these beliefs are systematically false.

The question is whether these two conditions can be jointly fulfilled. Let us focus on a particular normative judgment:

- 3) Lying is wrong.

What is one doing when one makes this judgment? Non-cognitivists say that one is expressing a conative (non-cognitive) state that motivates one not to lie, whereas cognitivists say that one is expressing a belief whose propositional content is that lying is wrong. If one claims that the judgment that lying is wrong is a non-cognitive state, then the question of its truth or falsity doesn't arise, and claiming that it is false makes no sense. If all normative judgments were classified as non-cognitive states, an error-theory thus would be unreachable. If one claims that the judgment that lying is wrong instead expresses a belief that lying is wrong, one must fulfill all the conditions required to make such a belief-ascription. Let us focus then on this judgment:

- 4) John believes that lying is wrong

What is involved in making this judgment? According to my implementation of the Kantian strategy for securing the objectivity of doxastic norms, ascriptions of belief require making normative judgments: to ascribe a belief one must judge, implicitly at least, that the mental state so classified is correct if and only if it is true. Judging that John believes that lying is wrong thus commits one to the normative judgment that:

- 5) John's belief that lying is wrong is correct if and only if lying is wrong.

How are we to understand this commitment to a norm of correctness? If we interpret it as a non-cognitive state, then, as before, the possibility of an error-theory cannot arise with respect to the normative judgment expressed by it. To judge that the belief that lying is wrong is correct if and only if lying is wrong would not itself be a belief, and thus would be incapable of being true or false. Suppose that the commitment to the norm of correctness is a belief whose content is that the belief one has ascribed has the normative property of being correct if and only if its content is true. Thus, ascribing the belief that lying is wrong would require believing the following normative proposition:

6) The belief that lying is wrong is correct if and only if lying is wrong.

If one were an error-theorist about normative judgments—believing amongst other things that there is no normative property of wrongness and thus no true propositions attributing this property to anything—one therefore would be committed to believing the following two inconsistent claims:

7) There are no true normative propositions.

8) John's belief that lying is wrong is incorrect.²⁷

The latter claim is itself a normative proposition that must be false if the former claim is true. And this case obviously generalizes to attributions of all normative beliefs. Attributing normative beliefs to others commits one to believing normative propositions oneself.

²⁷ The inconsistency can be made more transparent if we frame the two claims the error-theorist is committed to in terms of properties. Here are the two claims reformulated:

7*) There are no normative properties.

8*) John's belief that lying is wrong has the normative property of being incorrect.

Those whose metaphysical scruples cause them to recoil at talk of the existence of properties obviously will reject this formulation of the error-theorist's commitments.

The error-theorist thus appears to face a dilemma. He claims that all normative judgments attributing normative properties are false, that there are no normative truths that would make some of our normative judgments true.²⁸ He must decide whether normative judgments express beliefs or not. If he decides that they do not, then he must admit that normative statements, contrary to surface appearances, do not express propositions. If normative statements do not express propositions, then it is not the case that they are false, contrary to the error-theorist's central contention. But if the error-theorist decides that normative judgments express beliefs, and attributing a belief entails making a normative judgment, he cannot consistently claim that all normative statements are false.

Whether ascriptions of normative judgments are interpreted as expressions of beliefs or non-cognitive states, if the Kantian strategy for securing the objectivity of doxastic normativity succeeds, it appears that the conclusion that all normative judgments are false cannot be consistently reached. A purely detached external perspective on normative discourse is impossible. To attribute a normative judgment is already to take a normative stand.

If this argument is correct, the error-theory is inconsistent, and thus false. But does the argument entail that any of our normative judgments are actually true? No, because the argument doesn't show that our normative judgments even express beliefs capable of truth or falsity.

An error-theorist must attribute normative judgments and demonstrate that they are all false. The argument I described aims to show that fulfilling the first condition makes it impossible to satisfy the second condition; but, as I just pointed out, this tells us nothing about whether any of our normative judgments are true. Are there further conditions that an error-theorist must satisfy, and is there a way of moving from the claim that an error-theorist must satisfy these further conditions to the truth of at least one normative

²⁸ This formulation of an error-theory is intended to exclude judgments such as 'It is not the case that lying is wrong' that don't appear to attribute any normative property. 'It is not the case that lying is wrong' is not equivalent to 'It is permissible to lie'.

judgment? Presumably, in asserting that all normative judgments are false, the error-theorist is aiming to assert the truth. The error-theorist, it might be argued, thus takes himself to believe that all normative judgments are false, which means that he must take himself to have a mental attitude towards the proposition that all normative judgments are false that itself has the normative property of being correct if and only all normative judgments are false. This, though, is merely a ‘pragmatic’ contradiction— in order to sincerely assert that all normative judgments are false, one must take one normative proposition to be true; it does not follow that any normative judgments are true.

Of course, if it could be established that normative judgments express beliefs, then it would follow that for every normative judgment there would be a corresponding true normative proposition to the effect that the belief expressed by the judgment is correct if and only if its content is true. But establishing the conditions that need to be met in order to arrive at an error-theory at best tells us that if anyone believes an error-theory then at least one normative proposition is true; it would not tell us whether the antecedent of this conditional is true—whether anyone ever has believed an error-theory— and therefore would not tell us whether there are any true normative propositions.

But if the argument I have given is correct (and we accept it), it would remove the anxiety induced by the thought that all our normative beliefs are false because the universe contains no normative facts that could make them true. If my argument is correct, then it follows from the fact that someone has a normative belief that there is at least one normative fact. So any argument that attempts to demonstrate that all of our normative beliefs are false is bound to fail.

This result also has relevance for disputes about error-theories such as Mackie’s that limit themselves to claiming that a certain sub-domain of normative judgments, such as moral judgments, are systematically false. An error-theorist about moral judgments can consistently accept the following pair of judgments:

7) There are no true moral propositions.

8) John's belief that lying is wrong is incorrect.

The second claim is not itself a moral claim and thus does not conflict with the first one. It is not incoherent to believe an error-theory about morality. The question is how one can arrive at such a view. If the argument I just gave is correct, one cannot consistently argue for an error-theory about a specific sub-domain of normative judgments by focusing on a feature of those judgments that is shared by normative judgments in general. For example, an argument that moral judgments are systematically false because they involve attributions of normative properties, but normative properties being non-natural properties, don't exist, will apply equally to all judgments that attribute normative properties. If an error-theory about normative attributions in general is self-contradictory, this argument is defective; it cannot lead one to a wholesale denial of moral truth.

Mackie himself claimed that moral judgments were peculiar because of their intrinsically action-guiding character. This feature of moral judgments is what, according to him, is responsible for their being systematically false. His implicit argument is something like this:

Moral judgments, being a species of normative judgments, must be capable of motivating those to whom they apply if they are to be true. Moral judgments, apply to agents independently of their desires. Therefore, moral judgments, to be true, must be capable of motivating agents independently of their desires. But only desires can motivate. Therefore, moral judgments cannot be true. On the other hand, normative judgments whose application to an agent depends on his antecedent motives, such as instrumental judgments about what an agent ought to do given that he has a certain end, can motivate those to whom they apply, and thus are capable of being true.

Korsgaard, however, has argued that the normativity of such hypothetical imperatives depends on the correctness of categorical normative judgments.²⁹ If Korsgaard is correct, Mackie's argument thus generalizes to all normative judgments: unless there are some

²⁹ See Korsgaard (1997).

true categorical normative judgments, all normative judgments are false. A defender of Mackie might be tempted to claim that this just shows that his argument is more powerful than he initially thought. But if a fully general error-theory is self-contradictory, Korsgaard's argument amounts to a *reductio ad absurdum* of Mackie's position.

How satisfying these results are depends on what one hoped one would discover by doing metaethics. If the argument I just gave is correct, a fully general error-theory of normative judgments is incoherent. The argument thus has the potential to assuage the worry that our normative beliefs might be uniformly false. Even if we accepted it, I think a discomfort would remain, however. Notice that the argument shows that a global error-theory is false without actually engaging with the metaphysical and epistemological arguments error-theorists such as Mackie give. Even if we accept the argument, we still don't know how to fit normative properties into a scientifically respectable conception of reality, nor do we have an explanation of how we are able to detect the presence of these properties, given the kinds of capacities we believe ourselves to have. Thus, we are bound to feel a tension in our conception of reality: we believe there are normative properties but we also accept a scientific conception of reality—including that part of reality that includes ourselves—that seems to exclude them. While not leading us to a fully general error-theory of normative judgments, such an argument could still cause the kind of discomfort that any recognized instability in our beliefs causes us. Having been brought to this unstable position, we would seek escape. What options would we have?

Notice first that the more that the error-theorist's argument persuades us, the less we should be willing to apply the truth norm to our normative judgments. After all, the error-theorist's argument is intended to convince us that no norms are ever correctly applied, including the truth norm. There are two ways that we can withdraw our application of the truth norm to our normative judgments. One is to deny that normative discourse is truth-apt. This noncognitivist move would allow us to accept that our normative judgments do not conform to the norm of truth, not because our normative judgments are all false, but because normative judgments, including judgments endorsing or rejecting the truth norm itself, do not have truth-evaluable contents.

If normative judgments were expressions of non-cognitive states there would be no substantive question whether these judgments have or lack epistemic credentials. Because normative judgments would not express genuinely cognitive states, there would be no sense to be made of justifying these judgments by showing that their truth-conditions obtain. It would be open to us, though, to claim that showing that a normative judgment is inescapable is a way of justifying it in a different sense. As Kantian non-cognitivists, what would we mean by our assertion that a normative judgment is justified by being shown to be inescapable? We would be expressing our acceptance of the norm that favors any norm that one must accept in order to be an agent. Furthermore, we might claim that acceptance of this higher-order norm is itself implied by the acceptance of any normative judgment, or is part of the logic of agency.

On the other hand, we could accept that normative discourse is truth-apt, but nonetheless demur from applying the truth norm to our normative judgments. Instead, we might treat those judgments as useful fictions. As fictionalists about normative judgments we could accept the error-theorist's claim that all of our normative judgments are false, we would just deny that this makes them incorrect. It would be open to us to apply a different norm to normative judgments, however, which they might satisfy or fail to satisfy. For example, we might apply the norm of utility to our normative judgments, claiming that they are justified not in the epistemic sense of being supported by good evidence, but in the sense that making them is conducive to our overall happiness. Of course, applying the norm of utility to our normative judgments would itself consist in a false normative judgment, but we would determine whether to accept this judgment by whether making it maximizes utility, not by whether it is true. As Kantian fictionalists, all we would have to do to justify those normative judgments that are prerequisites for making any normative judgments at all is to demonstrate that there is a set of normative judgments that it would serve our interests to make. After all, if there is such a set, then it will include the judgments that the Kantian claims are inescapable for any creature capable of making any normative judgments at all.

The Kantian strategy puts a limit to the damage that an error-theorist can inflict, but it does not entirely remove the disquiet that an error-theorist's argument induces. The disquiet can no longer manifest itself in the worry that our normative practices are based on a mistake; nonetheless, it will continue to exist as a felt tension in our views about the nature of reality. This anxiety is bound to seek relief in some form of metaethics.

References

- Blackburn, Simon. (1998). *Ruling Passions*. Clarendon Press.
- Boghossian, Paul. (1990) 'The Status of Content,' *The Philosophical Review*, 99:2, 157-184.
- (2003). 'The Normativity of Content,' *Philosophical Issues* 13, 31-45.
- Boghossian, Paul, and Velleman, J. David. (1989). 'Colour as a Secondary Quality,' *Mind*, vol. 98, no. 389, January 1989, 81-103.
- Enoch, David. (2006). 'Agency, Schmagency: Why Normativity Won't Come from What is Constitutive of Action.'" *Philosophical Review*, 115:2, 169-198.
- Gibbard, Allan. (2003) *Thinking How to Live*. Harvard University Press.
- Korsgaard, Christine. (1996). *The Sources of Normativity*. Cambridge University Press.
- (1997). 'The Normativity of Instrumental Reason,' edited by Garrett Cullity and Berys Gaut. Oxford University Press, 215-54.
- (2003). 'Realism and Constructivism in Twentieth-Century Moral Philosophy,' *The Journal of Philosophical Research*, APA Centennial Supplement, 99-122.
- Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. Penguin.
- Rawls, John. (1971). *A Theory of Justice*. Harvard University Press.
- Shah, Nishi. (2003) 'How Truth Governs Belief,' *Philosophical Review*, 112:4, 447-82.
- Shah, Nishi, and Velleman, J. David. (2005). 'Doxastic Deliberation,' *Philosophical Review*, 114:4, 497-534.
- Shah, Nishi, and Hussain, Nadeem. (2006). 'Misunderstanding Metaethics: Korsgaard's Rejection of Realism,' *Oxford Studies in Metaethics 1*, edited by Russ Shafer-Landau, Oxford: Clarendon Press, 265-94.
- Smith, Michael. (1994) *The Moral Problem*. Blackwell Publishing.
- Street, Sharon. (2008) 'Constructivism about Reasons,' *Oxford Studies in Metaethics*, vol. 3, ed. Russ Shafer-Landau, Oxford: Clarendon Press.
- Stroud, Barry. (2000). *The Quest for Reality*. Oxford: Oxford University Press.

- Velleman, J. David. (2000a). 'The Possibility of Practical Reason,' reprinted in *The Possibility of Practical Reason*, Oxford University Press, 170-199.
- (2000b). 'Deciding to Decide,' reprinted in *The Possibility of Practical Reason*, 221-43.
- (2000c). 'The Aim of Belief,' in *The Possibility of Practical Reason*, 244-81.
- (2009). *How We Get Along*. Cambridge University Press.
- Wedgwood, Ralph. (2002). 'The Aim of Belief,' *Philosophical Perspectives* 16, 267-97.
- Williams, Bernard. (1995) 'Ethics and the Fabric of the World,' in *Making Sense of Humanity, and other Philosophical Papers 1982-1993*. Cambridge University Press. 172-181.