The Limits of Normative Detachment

Nishi Shah
Amherst College
Draft of 10/23/08

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”

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. 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. Moral statements attain objectivity, not

---

1 Previous drafts of this paper were discussed at the Mid-Atlantic Reading Group in Ethics (MARGE), a graduate seminar at NYU, the University of Pittsburgh, and the 2008 OSU/Maribor/Rijeka Conference in Dubrovnik on Valuing Agents. For additional discussion thanks to Jim Bell, Paul Boghossian, Michael Bratman, Ruth Chang, David Enoch, Matt Evans, Jyl Gentzler, James Harold, Paul Horwich, Nadeem Hussain, Casey Perin, Kieran Setiya, Matthew Silverstein, John Skorupski, Sharon Street, and David Velleman.


by accurately reflecting aspects of normative reality, but by expressing inescapable commitments of practical agency. 4

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.5 More specifically, one must apply the norm:

Believing that p is correct if and only if p.6

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.7 My use of the label

---

4 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. 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 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 is it 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.
7 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-
‘Kantian’ is meant to place these arguments in a certain tradition inspired by reading Kant, not to attribute these arguments to Kant himself.

I believe that there has been substantial confusion about the *metaethical* implications of the Kantian strategy. For example, both defenders and critics of Christine Korsgaard’s influential recent application of the Kantian strategy to ethical norms treat her position as an alternative to traditional cognitivist and non-cognitivist metaethical views.\(^8\) However, as far as her Kantian strategy is concerned, commitments to these norms may or may not express cognitive attitudes capable of truth or falsity, and even if they do express beliefs, her arguments do nothing to show that the content of those beliefs are true or likely to be true. All that Korsgaard’s Kantian strategy shows is that in some sense of ‘commitment’ not specified by her argument, one must be committed to these norms in order to be an agent. But the fact that one must apply a norm in order to be an agent, even if the attitude involved in applying it is interpreted as a belief with a truth-conditional content, does not imply anything about whether that truth-condition obtains or how we might come to know whether or not it obtains. The Kantian strategy thus does not seem to preclude the possibility that being an agent requires one to hold mistaken beliefs.

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

---

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 this kind of explanation as applied to doxastic normativity.

\(^8\) See Hussain and Shah (2006, ms) for a discussion of Korsgaard’s specific version of the Kantian strategy.
Williams describes. 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.

The Kantian strategy thus appears to be metaethically neutral. Whether this appearance of neutrality is veridical depends on whether the Kantian strategy really does 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 standards—but of being duped. If we are worried by Mackie’s argument, it is because

---

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.

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.
we are worried that ethics is an illusion, based on truth- evalu able judgments that couldn’t possibly be true, in the same way that witchcraft is a practice based on beliefs about the supernatural that couldn’t possibly be, or at least clearly aren’t, true. Being told that everyone is under this illusion, or even that it is impossible for anyone not to be under this illusion, 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

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 illusion. Is there a way of understanding the Kantian strategy as a method of arriving at normative truths? If not, might the Kantian strategy be used to undermine the threat that we are duped in our judgments about such fundamental norms in some other way? If the Kantian strategy successfully either provided a method of arriving at 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.

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 ground 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 Korsgaard is correct that to be an agent one must accept that humanity is valuable. From these claims it does not follow that humanity 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.
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 doxastic and practical norms 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 for arriving at normative truths, the Kantian strategy, if applied to both practical and doxastic norms, 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

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 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 arriving at normative truths?
The Kantian strategy, if it succeeds, accomplishes one very important task, which is to show that our fundamental normative judgments are deliberatively 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 true. This form of justification thus does not seem to shed any light on the *epistemic* credentials of the judgments that are so justified.

Our question 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 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, 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.

However, if normative judgments are about facts that are constructed from some set of privileged normative judgments, the Kantian strategy may be able to provide us with a method for arriving at normative truths. For example, if the truth-conditions for our actual normative judgments are constituted by the set of judgments that we would make in light of full empirical information, then the Kantian strategy, by showing that a certain class of judgments is inescapable for us, and thus must be part of any set of normative judgments that we could arrive at, would epistemically warrant those judgments. If normative facts are constituted by facts about the set of judgments we would arrive at 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 arrive at is a way of showing that that class of normative judgments is true.

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 normative judgments. 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.

---

13 ‘Full information’ accounts of the truth-conditions 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-conditions 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.
Is there a way though of understanding the Kantian strategy as itself an argument for constructivism? 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 independent of our normative judgments, as the error-theorist supposes, the Kantian strategy would not provide such a demonstration. Does this mean that the Kantian needs an independent argument for the claim that normative truths are not independent of normative judgments to defeat the error-theorist? This depends on what attitude we are allowed to take towards the Kantian strategy antecedently to engaging in metaphysical inquiry. If we are *prima facie* entitled to treat the Kantian strategy as a procedure for arriving at 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 can see no reason to accept that our metaphysical commitments must be arrived at independently of our epistemological views. 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.\(^\text{14}\) We thus don’t need an independent reason for thinking that normative facts are constructed out of normative judgments; 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.

On this interpretation of the Kantian strategy, its epistemic credentials thus 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

\(^{14}\) Or, to put things in more clearly epistemic terms, by treating the Kantian strategy as a method of epistemic justification, we commit ourselves to whatever ontology best explains how it could yield normative knowledge. Given that we are assuming cognitivism here, it is difficult to see why we should abjure from applying this description to the upshot of a sound deployment of this strategy.
must constitute a determinate view of normative truth. As I will now argue, this is something we have very good reason to doubt.

Constructivism is a thesis of metaphysical priority: normative properties are grounded in or dependent on normative judgments. For example, 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. This general class of views can be represented schematically in the following way:

**Metaethical Constructivism:**

The property of being F, where F is a normative predicate, is such that for any x, if x is F, then x is F because of S’s judgment 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..

This is a schema for a determinate view only if there is a specification of the type of judgment out of which the property of being 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 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.\(^{15}\)

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

\[^{15}\] 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.
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 beliefs. To say that the relevant facts are constituted out of our responses 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. It might be thought that I am arguing that the existence of beliefs with normative content are asymmetrically dependent on the existence of normative facts, and therefore that normative facts cannot be constituted out of normative beliefs.

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 without any 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 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.\textsuperscript{16} It thus appears that the constructivist has no way to specify the judgments that he believes constitute the metaphysical ground of normative truths. At the very least, he does not have the resources to distinguish between F properties and G properties, if both properties are constructed out of judgments. For example, if the property of moral rightness is judgment-constituted, presumably so too is the property of moral wrongness. But if the contents of the beliefs that constitute these properties are indeterminate, there will be no facts about which judgments are rightness judgments and which are wrongness judgments, and thus no facts about which acts are right and which are wrong.

One way to try to avoid this problem is to analyze the normative properties not in terms of judgments attributing those very properties, but in terms of some other judgments attributing different normative properties. The view would have the following structure:

\textbf{Metaethical Constructivism*}:  
The property of being F, 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 judgments, 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 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 these new judgments or responses 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 truths to normative judgments. The truth of such bi-conditionals entails only

\textsuperscript{16} Compare Boghossian and Velleman’s objection to a similar account of color facts, (1989), 89-90.
that the relevant normative truths are epistemically accessible to the relevant agents; they
do not establish that those normative truths are constructed out of the corresponding
normative judgments. In order for normative properties to be constituted by normative
judgments, normative judgments must be metaphysically prior to them. But if there is no
way to identify any normative judgments without mentioning normative properties, this
metaphysical priority cannot be established.

Let us assume that there is a way for the constructivist to specify determinate truth-
conditions for the relevant normative judgments.\(^{17}\) 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 property of being 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 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.\(^{18}\) If this thesis is true, for whatever reason,
it is difficult to understand what metaethical constructivism could be.

\(^{17}\) See Velleman (2000b) for an attempt to solve this problem as it arises in the context of specifying the
content of joint agreements.

\(^{18}\) 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.
Metaethical constructivism, remember, is supposed to be a fully general account of normative truth, therefore it needs an interpretation of all normative judgments, whether those judgments express the acceptance of doxastic or practical norms. 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 property 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 property. Otherwise, for all we know, the property may exist in an independent normative realm to which we have no epistemic access. And if it does, so too does 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).

To complete this account, we now need a constructivist account of S’s belief that R believes that \( p \), since beliefs are themselves constituted by normative properties according to the constructivist. (Remember, the constructivist is not an expressivist who denies that normative predicates express properties.) The truth-condition of S’s belief that R believes that \( p \) therefore must be constructed out of further beliefs if a fully general constructivism about normativity is correct. The account thus has the following structure:

S’s belief that \( p \) is constituted out of T’s belief that S believes that \( p \), which is itself constituted out of U’s belief that T believes that S believes that \( p \), etc...

We again end up with a regress, this time with beliefs being constructed out of further beliefs, which in turn must be constituted out of further beliefs, and so on *ad infinitum*. 
The prospects for metaethical constructivism look bleak. The constructivist claims that normative properties are constituted out of normative judgments. 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 properties or be drawn into a vicious regress.\textsuperscript{19}

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 normativity 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 shown to 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 the falsity of some Kantian claims that certain normative judgments are constitutive of being an agent, then the Kantian can refute the error-theorist by demonstrating the inescapability of those 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

\textsuperscript{19} My thinking about constructivism has benefited immensely from reading Sharon Street’s paper ‘Constructivism about Reasons,’ which of course is not to say that she would agree with what I have said about it.
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 must be accepted 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.\footnote{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.}

What must one do to arrive at a fully general error-theory of normativity? It is not sufficient that one assert that propositions attributing normative properties 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 arrive at an error-theory about normativity 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 normativity—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.\(^{21}\)

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.

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.\(^{22}\) 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

---

\(^{21}\) 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.

\(^{22}\) 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’.
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, does it entail that any of our normative judgments are actually true? No. First of all, the argument doesn’t show that our normative judgments even express beliefs capable of truth or falsity, and even if we assume that they do, all it shows is that one cannot believe one thing without believing another: I cannot believe that John believes that lying is wrong without myself holding a normative belief, namely that John’s belief is correct if and only if lying is wrong. Nothing follows as to the truth or falsity of either belief. If constructivism were correct the truth of a normative proposition would follow from it’s being believed under certain conditions. Unfortunately, as we saw, constructivism cannot be made fully coherent. Absent such an account, we have no license to draw such an inference.

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 the error-theorist’s central claim is true or false. 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 proposition? 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, again, 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.

If we could show not merely that the error-theorist must take himself to believe that all normative judgments are false, but that he must actually believe that all normative judgments are false, would this establish that there is at least one true normative proposition? No, what would be established is that it is impossible for an error-theory to be true and believed— in order to believe an error-theory, one must satisfy one of the conditions that the theory itself says cannot be satisfied, namely believing something; from this claim it does not follow that any normative judgments are true—it does not follow that anyone ever does satisfy this condition, that anyone ever does believe anything. 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 is correct (and we accept it), it would remove the anxiety induced by the thought that we might discover that all our normative beliefs are false. If it is correct, we cannot consistently reach that result. No argument could ever demonstrate that everyone’s normative judgments are all false, because any such argument would have to attribute normative beliefs to us, and attributing normative beliefs already requires committing oneself to some normative truths. Thus, a fully general error-theory
cannot be reached, because any argument for such a conclusion will commit itself to at least one normative truth.

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 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. Normative judgments whose application to an agent depends on his antecedent motives, such as judgments about what
an agent ought to do to fulfill 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. If Korsgaard is correct, Mackie’s argument thus generalizes to all normative judgments: unless there are some 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 (this is certainly what I used to think). But if a fully general error-theory is self-contradictory, Korsgaard’s argument amounts to a reductio of Mackie’s position.

How satisfying these results are depends on what one hoped one would discover by doing metaethics. The argument I have given that an error-theory of normative judgments in general is unreachable does not shed light on the metaphysics or the epistemology of those judgments: we still don’t know, as far as the argument goes, whether normative judgments express propositions about normative facts, and if so, what the nature of these facts is and how we can come to know them. But these questions, though not resolved, surely will feel less urgent (possibly idle?) once the threat that our normative judgments might all be ‘unmasked’ as false has been laid to rest.

I think a discomfort would remain, however. If the argument I just gave is correct, we cannot get ourselves in the position of consistently accepting a fully general error-theory of normativity. But we might feel the force of a Mackie-style argument nonetheless. After all, the argument against the error-theorist does not tell us where his argument goes wrong, and it certainly doesn’t show that our normative judgments do successfully pick out normative aspects of reality nor does it tell us by what means we come to track them if they do exist. A Mackie-style argument could lead us to recognize a tension in our views about the nature of reality and the extent of our epistemic access to it. 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 in which we feel that our normative judgments are defective even though we recognize that even to think this
requires making some normative judgments, we would seek escape. What options would we have?

Notice first that the more that the error-theorist’s argument works on 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 persuade 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 option 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, cannot be assessed for their truth-value in the first place.

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 accept 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 by treating those judgments as useful fictions. As fictionalists 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 by refusing to apply the norm of truth to them. 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.

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


——— (ms) ‘Metaethics and its Discontents: A Case Study of Korsgaard.’


Street, Sharon. (forthcoming) ‘Constructivism about Reasons.’


