

## Durham Research Online

---

### Deposited in DRO:

28 April 2020

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Cowie, Christopher (2020) 'Do constitutive norms on belief explain Moore's Paradox?', *Philosophical studies.*, 177 (6). pp. 1685-1702.

### Further information on publisher's website:

<https://doi.org/10.1007/s11098-019-01280-6>

### Publisher's copyright statement:

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.



# Do constitutive norms on belief explain Moore's Paradox?

Christopher Cowie<sup>1</sup>

Published online: 25 March 2019  
© The Author(s) 2019

**Abstract** In this article I assess the prospects for a particular kind of resolution to Moore's Paradox. It is that Moore's Paradox is explained by the existence of a constitutive norm on belief. I focus on a constitutive norm that relates belief to knowledge. I develop this explanation. I then present a challenge to it. Norm-based explanations of Moore's Paradox must appeal to a 'linking principle' that explains what is wrong with violating the constitutive norm. But it is difficult to articulate a plausible candidate linking-principle. I show this by canvassing the obvious candidates and articulating the problems with each.

**Keywords** Moore's Paradox · Constitutive norms · Belief

We can seemingly make judgments of the form:  $p$ , but I don't believe  $p$ . But there would be something problematic about doing so. What exactly is wrong with judgments of this form and what explains it? This is Moore's Paradox in its simplest form. It has been thought to reveal something interesting about the nature of belief, assertion or the nature of subjective experience itself. In this article I assess the prospects for a particular kind of resolution. It is that Moore's Paradox is explained by the existence of a *constitutive norm on belief*. This kind of solution has been popular in the recent literature.<sup>1</sup> I will label and understand it as follows:

---

<sup>1</sup> See Huemer (2007). For discussion see Littlejohn (2010) and de Almeida (2001). For some critical discussion see Lo Presti (2014). A closely related view is discussed in Baldwin (2007). For recent comprehensive discussion of Moore's Paradox—including specific reference to norm-based solutions—see Williams (2015a, b) and Green and Williams (2007).

---

✉ Christopher Cowie  
christopher.d.cowie@durham.ac.uk

<sup>1</sup> Department of Philosophy, University of Durham, Durham, UK

*The Norm-Based Explanation* Moore's Paradox is explained by the existence of a constitutive norm on belief, which is violated by a Moore-paradoxical judgment.

The explanation will vary depending on the form that the constitutive norm takes. On most accounts it is either truth-based or knowledge-based. This gives rise to the following two variants.

*The Truth-Norm Based Explanation* Moore's Paradox is explained by a constitutive norm relating belief to *truth*, which is violated by a Moore-paradoxical judgment.

*The Knowledge-Norm Based Explanation* Moore's Paradox is explained by a constitutive norm relating belief to *knowledge*, which is violated by a Moore-paradoxical judgment.

This kind of solution—whether truth-based or knowledge-based—is attractive for a number of reasons. Firstly, norm-based explanations entail that in making a Moore-paradoxical judgment one has done something *wrong* or *incorrect* or something that one *shouldn't*. This is intuitively appealing. Surely, one has done something wrong, or incorrect, or that one shouldn't (in some sense at least) in making a Moore-paradoxical judgment. And yet not all purported solutions to Moore's Paradox capture this. Suppose, for example, that one were to argue that it is simply impossible, as a matter of psychological fact, to make Moore-paradoxical judgments.<sup>2</sup> This would fail to speak to the fact that there is something wrong or incorrect about so believing. Norm-based solutions are attractive in part because they build this feature—the wrongness, or incorrectness—into the solution from the very start.

Secondly, the kind of norm-based explanation that I am interested in doesn't simply render Moore-paradoxical judgments wrong or incorrect relative to a common-or-garden epistemic norm. It renders them wrong relative to a norm that is constitutive of belief itself. This is important. It is natural to think that in making a Moore-paradoxical judgment one has done something that sits uncomfortably with the nature of belief or believing itself and not merely that one has made an ordinary epistemic mistake as one might if one believed an evidentially unsupported proposition.<sup>3</sup> The norm-based explanations that I am interested in—explanations that appeal to constitutive norms on belief—build this feature into the solution from the very start. They make Moore-paradoxical judgments not only wrong, but wrong in virtue of what it is for one to believe something at all.

Thirdly, norm-based explanations, and in particular knowledge-norm based explanations explain a range of closely related phenomena. This is to the fore in Huemer's defence of the knowledge-norm based solution.<sup>4</sup> Huemer claims that

<sup>2</sup> Explanations in terms of the self-intimating nature of belief—i.e. such that if one believes *p*, then one believes that one believes *p*—are arguably of this form. See e.g. Shoemaker (1996).

<sup>3</sup> In the literature Moore's paradox is sometimes explained by appeal to simple epistemic norms on rational credibility e.g. Douven (2009).

<sup>4</sup> Huemer, *ibid.*

Moore-paradoxical judgments of the form 'p, but I don't believe p' could be well-explained by the self-intimating nature of belief: that if one believes p, then one believes that one believes p. But he notes that this solution wouldn't explain what is wrong with a range of closely-related judgments, such as judgments of the form: 'p, but I don't know that p' and 'p, but I don't have any evidence for p'. By contrast the knowledge-norm based solution (which will be my focus) does explain what's wrong with this range of judgments. They all violate the knowledge-norm on belief. As such the knowledge-norm based explanation provides a unified explanation of a wide range of closely related phenomena.

Fourthly and finally, the norm-based explanation fits nicely with a wider pattern of explanation in epistemology and philosophy of mind. Quite independently of Moore's Paradox, there has been a surge of recent interest in constitutive norms on belief in part due to the role that they can play in explaining phenomena in mind and language.<sup>5</sup> Part of the motivation for taking norm-based solution to Moore's Paradox seriously is that they fit with this broader program. This is especially true of knowledge-norm based explanations given the success of the knowledge-first program of explanation of phenomena in mind and language.<sup>6</sup>

Given these four motivations norm-based explanations of Moore's Paradox—and particularly knowledge-norm based explanations—should be taken seriously. However while there has been some recent back-and-forth between defenders of truth-norm based explanations and knowledge-norm based explanations respectively (which is better?), there has been little general, critical discussion of norm-based explanations per se.<sup>7</sup> This is something of an oversight. There are systematic issues that norm-based solutions face. My aim in this article is to articulate one such issue. I argue that norm-based explanations must be supplemented by what I refer to as a *linking principle*: a principle that explains how and why the violation of a constitutive norm on belief is problematic. I will argue that (1) some such principle is required if norm-based explanations are to work, (2) this point has been largely ignored in the literature, and (3) it is hard to find a satisfactory principle that doesn't come with substantial extra costs. I do my best on behalf of norm-based explanations to find such a principle. I do not succeed. My conclusion is that optimism for norm-based explanations should be tempered.

The plan for the article is as follows. I begin in Sect. 1 by making some quick remarks about the view that there are constitutive norms on belief. In Sect. 2 I make some preliminary remarks about the application of this strategy to the resolution of Moore's Paradox. I identify some desiderata for a solution. In Sects. 3–4 I present my main argument: attempts to resolve Moore's Paradox by appeal to knowledge-based constitutive norms on belief must make use of a linking-principle, and it is hard to see what this should be. In Sect. 5 I conclude.

<sup>3</sup> In the literature Moore's paradox is sometimes explained by appeal to simple epistemic norms on rational credibility e.g. Douven (2009).

<sup>4</sup> Huemer, *ibid.*

<sup>5</sup> See e.g. Shah (2003) and Shah and Velleman (2005) who explain the transparency of belief in terms of a constitutive norm on (the concept of) belief.

<sup>6</sup> Most closely associated with Williamson (2000). For development see Carter et al. (2017).

<sup>7</sup> See Huemer (*ibid.*) and Littlejohn (*ibid.*).

## 1 Constitutive norms on belief: preliminary remarks

The claim that there are constitutive norms on belief can take many different forms. At its most general we can represent it as follows:

*Normativism*: It is constitutive of belief that: beliefs or believers stand in some normative relation to truth or knowledge.<sup>8</sup>

By ‘constitutive’ I understand ‘at least part of what it is to be’. So the normativist view is that it is at least part of what it is for a state to be one of belief that: beliefs or believers stand in some normative relation to truth or knowledge. The normativist view has been cashed out in lots of different ways. The following versions exist—there or thereabouts—in the literature:

- (1) It is constitutive of belief that: one ought to believe  $p$  iff  $p$ .
- (2) It is constitutive of belief that: one ought to believe  $p$  only if  $p$ .
- (3) It is constitutive of belief that: one may believe  $p$  iff  $p$ .
- (4) It is constitutive of belief that: it would be good to believe  $p$  iff  $p$ .
- (5) It is constitutive of belief that: it would be fitting to believe  $p$  iff  $p$ .
- (6) It is constitutive of belief that: one ought to believe  $p$  only if one knows  $p$ .<sup>9</sup>

I won’t run through the strengths and weakness of the various formulations here. I simply note three important points.

Firstly and most obviously, (1)–(5) are all versions of a *truth-norm* on belief. (6) is a *knowledge-norm*. My argumentative focus will be solely on knowledge-norms. Secondly, these formulations are actually ambiguous given different available readings of the scope of the normative or evaluative operator. Each can be read as taking either narrow-scope or wide-scope. So (1) could be read either as a narrow-scope requirement:

(1—Wide) It is constitutive of belief that: One ought to (believe  $p$ ) iff  $p$ .

Or as a wide-scope requirement:

(1—Narrow) It is constitutive of belief that: One ought to (believe  $p$  iff  $p$ ).

These are different principles that require different things of believers.<sup>10</sup> Similar comments apply to (2)–(6). I shall flag if at any point below this becomes relevant in a sense that could be contentious or confusing. Thirdly, it is commonplace to formulate the normative relation in terms of *correctness*. So, for example, one might see any of the following:

<sup>8</sup> McHugh and Whiting (Ibid.). For an influential defence see Wedgwood (2007). For a nice contemporary summary and defence see Engel (2014).

<sup>9</sup> Variants on (1)–(2) are defended—sometimes in terms of ‘correctness’ (see below)—by Shah (2003), Wedgwood (2002, 2007) and Boghossian (2003). (3) is defended in Whiting (2010) and (4) in McHugh (2012). (5) is discussed by McHugh and Way (2016) and Howard (forthcoming). (6) is inspired by Williamson (2000) and discussed in Huemer (2007).

<sup>10</sup> For a clear discussion see Bykvist and Hattiangadi (2007).

- (7) It is constitutive of belief that: it is correct to believe  $p$  iff  $p$ .  
 (8) It is constitutive of belief that: it is correct to believe  $p$  only if  $p$ .  
 (9) It is constitutive of belief that: it is correct to believe  $p$  only if one knows  $p$ .

'Correctness' can be used in different ways. It is sometimes used to indicate that the relation between belief and truth or knowledge—whatever it turns out to be—is a mere rule or requirement and is not genuinely normative any more than the rules of chess are. I have in mind here the kind of distinction appealed to by John Broome (amongst others) when he writes:

In one sense 'normative' simply means to do with norms, rules or correctness. Any source of requirements is normative in this sense.... But I do not use 'normative' in that sense. In my sense it means to do with ought or reasons. Given a rule or requirement we can ask whether you ought to follow it, or whether you have reason to do so.<sup>11</sup>

We could understand the constitutive norm on belief—whatever it may be—as either a mere correctness-norm like the rules of chess or as normative in Broome's sense. For the most part in this article I will simply work with (1)–(6)—and really just (6)—without specifying whether the property referred to in it is genuinely normative or is a 'mere' correctness norm. Later in the article, where this becomes important, I will return to it. (Basically, the point will be that if the constitutive norm is one of mere correctness it may be too weak to do the explanatory work required of it).

## 2 Resolving Moore' paradox: preliminary remarks

In this brief section I make some further preliminary remarks about Moore's Paradox itself. I begin by highlighting aspects of Moore's Paradox that I won't discuss (1–3) before drawing attention to one important desideratum for its resolution (4).

### 1. Varieties of Moore's paradox.

Moore's Paradox in the narrowest sense concerns judgments of the following forms (I use square brackets to indicate contents of judgments throughout):

*MP*: [ $p$ , but I don't believe  $p$ ].

*MP (Commissive)*: [ $p$ , but I believe  $\sim p$ ].

There are however a range of sentences that can seem similarly peculiar. As I mentioned above defenders of the knowledge-norm based explanation have claimed to be able to account for some of them. My focus however will be solely on *MP*.

### 2. Speech or thought?

<sup>11</sup> Broome (2007, p. 162) quoted from Olson (2014, p. 119).

Explanations of Moore's Paradox sometimes proceed at the level of thought, sometimes at the level of speech. Understood at the level of thought the challenge is to explain what's wrong with *believing* an instance of MP. Understood at the level of speech the challenge is to explain what's wrong with *asserting* an instance of MP. Different explanations of Moore's Paradox focus in different ways. My focus is on Moore's Paradox at the level of thought. It has been argued that explaining the paradox at the level of thought is more fundamental: that explanations at the level of speech are parasitic on this.<sup>12</sup> I think that this is probably right, but I won't presuppose it here.

### 3. Further subtleties.

The literature on Moore's Paradox is now extremely rich. There are many important issues that I bypass in this article given my aims. These include the significance of *conscious* thought in Moore's Paradox, accepting instances of MP diachronically as opposed to synchronically, and 'solutions' based on both the nature of evidence, justification and probability. In bypassing these I don't think that I am thereby being uncharitable to my opponents.<sup>13</sup>

### 4. A desideratum: the right kind of explanation.

I claimed above that there is 'something problematic' about believing an instance of MP. But this leaves a lot of room for interpretation. We should be more precise. What, exactly, is the problem with believing an instance of MP? This is contested. One should not stipulate an answer without care. Nevertheless it is important for my subsequent arguments that we can at least rule out some options as *too weak*. The way in which one goes wrong if one believes an instance of MP is not simply like the way in which one goes wrong when one believes a falsehood, or when—as in Gettier cases—one believes a truth accidentally, or by luck. It is stronger than this. Nor is it merely like the way in which one goes wrong when one believes a necessary falsehood such as a mathematical falsehood. This is also too weak. For the most part—including the next section, Sect. 3—these negative claims are all that my subsequent arguments will require. But I shall venture a more positive claim here nonetheless.

As I understand it the way in which one goes wrong when one believes an instance of MP is much more like the way in which one goes wrong when one holds blatantly inconsistent attitudes, for example when one both believes *p* and believes not-*p*. Obviously, the way in which one goes wrong when one believes an instance of MP cannot be *exactly the same as* the way in which one goes wrong when one holds blatantly inconsistent attitudes; beliefs in *p* and in not-*p* cannot simultaneously be true, but *p* and a belief in not-*p* can simultaneously be true (that's the whole problem). But this is, I think, the kind of defect that purported explanations of MP

<sup>12</sup> See e.g. Huemer (2010, p. 144) although Williams (2013) questions whether belief-based explanations of Moore's Paradox at the level of thought straightforwardly translate to explanations of Moore's Paradox at the level of assertion. For a full discussion of speech-based forms see Williams (2015b).

<sup>13</sup> Important references with respect to these issues include but are certainly not limited to Douven (2006), Sorensen (2000) and Smithies (2012). For an excellent summary see Williams (2015a).

should aim to approximate. They should explain why beliefs in instances of MP are problematic in a way that comes as close as possible to the way in which blatantly inconsistent attitudes are problematic. This is most naturally thought of as a form of *irrationality*.

So my tentative positive working-hypothesis is that believing an instance of MP is irrational in an approximation of the sense that accepting blatantly inconsistent attitudes is. Or as I shall say for short throughout, it is 'irrational in the relevant sense'. In any case, my arguments of the next section rely only on the weaker negative claim that the way in which one goes wrong in believing an instance of MP is not the same as the way in which one goes wrong by having a false or unjustified belief, even in a necessary truth. I shall have more to say about (ir)rationality in Sect. 4.

### 5. My argument.

With these necessarily brief remarks in place, we are now in position to look at norm-based explanations of Moore's Paradox. My main focus is on the knowledge-norm based explanation. This is my focus because it is the best or strongest candidate for a norm-based explanation of Moore's Paradox. As I hope will be clear to the reader though, the basic point of my argument would apply against truth-norm based explanations of Moore's Paradox too. My argument is that *any* norm-based explanation of Moore's paradox—whether truth-based or knowledge-based—will have to appeal to a *linking principle*: a principle that takes us from one's having violated a constitutive norm on belief to one's having done something irrational in the appropriate sense. I claim that articulating this principle is surprisingly problematic. I demonstrate this in the next two sections.

## 3 The knowledge-norm based explanation

Consider the knowledge-based norm on belief:

- (6) It is constitutive of belief that: one ought to believe  $p$  only if one knows  $p$ .

Can this be used to explain Moore's Paradox? I'm not aware of any really clear accounts in the literature of how such an explanation might go.<sup>14</sup> The following is my attempt at a 'best reconstruction'. The explanation has two steps. In the first step, we show that one cannot *know* an instance of a Moore-paradoxical sentence. This doesn't require appeal to any constitutive norms on belief. In the second step one appeals to a knowledge-norm on belief to explain what's wrong with accepting an instance of MP. It goes roughly as follows:

*Step 1:* In order to know an instance of MP one would have to know [ $p$ , and I don't believe  $p$ ]. If one knows a conjunction, then one knows each of its conjuncts. So in order to know an instance of MP one would have to know [ $p$ ] and know [I don't believe  $p$ ]. If one knows [ $p$ ], then one believes [ $p$ ]. So in order to know an

<sup>14</sup> The closest is Huemer (2007).

instance of MP one would have to believe[p] and know [I don't believe p]. But this is impossible. One cannot believe [p] whilst—as the factivity of knowledge entails—one simultaneously doesn't believe p. So one cannot know an instance of MP.

*Step 2:* Now add to this a constitutive norm on belief according to which: one ought not to believe p unless one knows p. In believing an instance of MP one violates that norm. One does so because that norm prohibits believing something one doesn't know. And Step 1 has shown us that you can't know an instance of MP.

We can represent the whole argument schematically. 'K[p]' represents that some agent knows that p, 'B[p]' represents that some agent believes that p, 'O[p]' represents that it ought to be the case that p.

#### Step 1

1. If  $K[p \& \sim B[p]]$ , then  $K[p] \& K[\sim B[p]]$  (*knowledge distributes across conjunction*).
2. If  $K[p]$ , then  $B[p]$  (*knowledge entails belief*).
3. (1,2) If  $K[p \& \sim B[p]]$ , then  $B[p] \& K[\sim B[p]]$ .
4. If  $K[\sim B[p]]$ , then  $\sim B[p]$  (*factivity of knowledge*).
5. (3,4) If  $K[p \& \sim B[p]]$ , then  $B[p] \& \sim B[p]$ .
6.  $\sim [B[p] \& \sim B[p]]$  (*non-contradiction*).
7. (5,6)  $\sim K[p \& \sim B[p]]$ .

#### Step 2

8.  $O[\text{If } \sim K[p] \text{ then } \sim B[p]]$  (*knowledge-norm on belief*).
9. (7, 8)  $O[\sim B[p \& \sim B[p]]]$ .<sup>15</sup>

This explains what goes wrong when one believes an instance of MP. One violates the knowledge norm on belief. Can this resolve Moore's Paradox?

Not as it stands. More work—another 'step'—is needed. Violating norms is something that most of us do a lot. I might violate a moral norm or a norm of etiquette. But I wouldn't thereby have done something irrational in the sense that it is irrational to believe an instance of MP. So why does a violation of the knowledge-norm make it irrational in the relevant sense to believe an instance of MP? We need an argument. More specifically, what we need is a *linking-principle*: a principle that takes us from the violation of the knowledge-norm on belief that occurs when believing an instance of a Moore-paradoxical sentence, to the problematic nature of so doing. What could that linking principle be? In this section I run through some candidates. Some are workable but none are fully satisfactory.

The first candidate linking principle states that it is irrational in the relevant sense—that is in a sense that approximates the irrationality of blatantly inconsistent

<sup>15</sup> For ease, premise 8 states the knowledge-norm on belief in the form that: it ought to be the case that [if one doesn't know p, then one doesn't believe p]. The move from 7 and 8 to 9 goes via an instance of this principle in which the content of the knowledge or belief is  $[p \& \sim B[p]]$ . Schematically, this could be represented by the interim premise:  $O[\text{If } \sim K[p \& \sim B[p]] \text{ then } \sim B[p \& \sim B[p]]]$ .

attitudes—to violate a *constitutive norm* on a practice, whilst engaging in that practice. Roughly:

**LP1:** It is irrational in the relevant sense to engage in a practice whilst violating the constitutive norm on that practice.

If combined with the knowledge-norm, LP1 uncontroversially delivers the result that believing an instance of MP is irrational in the relevant sense. Just to be clear, the argument is:

1. One cannot know an instance of MP.
2. It is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).
3. It is irrational in the relevant sense to engage in a practice whilst violating the constitutive norm on that practice (*LP1*).
4. (1–3) It is irrational in the relevant sense to believe an instance of MP.

Despite its promise this is obviously a bad argument. The linking principle LP1 is too strong. Its truth, combined with the knowledge-norm, entails that having beliefs that don't constitute knowledge is irrational in the relevant sense; that is, in some approximation of the sense in which having blatantly inconsistent attitudes is irrational. But clearly, this isn't true. Having a false or unjustified belief isn't irrational in anything like the sense that having blatantly inconsistent attitudes is. So LP1 is clearly too strong to be true.

There are a number of different ways of reformulating LP1—or tinkering with the other premises of the argument 1–4 above—that make might it look a bit more plausible. I will run through some of them below. Whist some are a definite improvement on LP1, none is fully satisfactory. The first proposal is to weaken LP1 to:

**LP2:** It is irrational in the relevant sense to engage in a practice whilst doing something that *necessarily* violates the constitutive norm on that practice.

Just to be clear, the argument is:

1. One cannot know an instance of MP.
2. It is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).
3. It is irrational in the relevant sense to engage in a practice whilst doing something that necessarily violates the constitutive norm on that practice. (*LP2*).
4. (1–3) It is irrational in the relevant sense to believe an instance of MP.

LP2 is better than LP1. LP2 doesn't entail, as LP1 does, that *all* beliefs that don't constitute knowledge are irrational in the relevant sense. But LP2 is still clearly too strong to be true. LP2 entails that it is irrational in the relevant sense to believe a necessary falsehood. This is false. Consider a false mathematical belief. Having such a belief is not irrational in the relevant sense: it is not irrational, as believing an instance of MP is, in a sense that approximates belief in blatantly contradictory

attitudes. This is sufficient to warrant rejecting LP2. It is also worth noting that LP2 faces a second more fundamental challenge (a challenge that could equally have been applied to LP1). LP2 is not a basic or primitive truth. In light of this we should ask why (and how) is it irrational in the relevant sense to engage in a practice whilst doing something that *necessarily* violates the constitutive norm on that practice. If LP2 is true, there must be some more basic or primitive answer. I'll discuss some possible answers to this later (when I discuss *rational requirements*—the obvious source of such grounds). But for now I simply leave it as an issue that would require a resolution.<sup>16</sup>

Consider an alternative modification of LP1. We could modify LP1 to:

**LP3:** It is irrational in the relevant sense to engage in a practice whilst doing something that *one believes* (or is committed to believing) violates the constitutive norm on that practice.<sup>17</sup>

So the argument we are considering is:

1. One cannot know an instance of MP.
2. It is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).
3. It is irrational in the relevant sense to engage in a practice whilst doing something that *one believes* (or is committed to believing) violates the constitutive norm on that practice. (LP3).
4. (1–3) It is irrational in the relevant sense to believe an instance of MP.

In one sense LP3 looks better than LP1 or LP2. LP1 entails that beliefs that don't constitute knowledge are irrational in the relevant sense (i.e. a sense that approximates the irrationality of blatantly inconsistent beliefs). LP2 entails that beliefs in necessary falsehoods are irrational in the relevant sense. These are problematic entailments. LP3 avoids both.

Despite this LP3 faces a bigger problem than either LP1 or LP2. The above argument in which it figures is invalid. 4 does not follow from 1–3. This is because its second premise is about *what one ought to believe*, but its third premise is about *what one believes about what one ought to believe*. We could get around this by modifying the second premise. We could make it about what one *believes about* the knowledge-norm. For example:

2\*. *One believes or is committed to believing that* it is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).

This would make the argument valid (assuming premise 1 is amended accordingly). But it is independently problematic. Firstly, 2\* neither entails the knowledge-norm, nor is 2\* itself entailed by the knowledge-norm. So even if we accepted the above

<sup>16</sup> One might argue that it doesn't require a resolution. However I take it to be a prima facie problematic feature of a proposed solution to a problem if it relies on a clearly non-basic principle for which there is no obvious candidate for a more basic principle that could explain it.

<sup>17</sup> I shall assume that there is a sensible way of understanding 'committed to' here that a defender of LP3 could make use of.

argument, we would need to see a lot more to be able to link it up with the knowledge-norm. Secondly, 2\* is not obviously true either. Amongst the small number of people who have views about the nature of belief many do not believe that it is constitutive of believing a proposition that one ought to do so only if one knows it. It follows from 2\* that these people don't do anything irrational in the relevant sense in believing an instance of MP. This is clearly ridiculous. Now of course one could claim in defence of 2\* that the aforementioned philosophers and psychologists are mistaken and that they, along with everyone else, are, contrary to their own judgment, *committed* to believing that it is constitutive of believing a proposition that one ought to do so only if one knows it. But anyone who claims this faces an obvious burden of proof.

It is worth saying a little bit more about this. Perhaps the most obvious way for a defender of the knowledge-norm to defend 2\* is to claim that the knowledge-norm on belief is true not merely of belief itself, but of *the concept* of belief. That is to say, they would need to claim that it is a conceptual truth about belief that: one ought to believe p only if one knows p. This would arguably make sense of 2\* in a way that would link it up to the knowledge-norm itself. If it were true, then anyone who is competent with the concept of belief would 'be conceptually committed' to believing that they ought to believe p only if they know p. And if this were true, then a defender of the knowledge-norm explanation *could* appeal to LP3 after all.

This strategy is not a total non-starter.<sup>18</sup> But it is problematic. Claims about 'conceptual truths' are very strong and the worry is that those who make them unjustly accuse anyone with non-standard views of being 'conceptually incompetent' when, clearly, they are not.<sup>19</sup> This generates an acute form of the worry with 2\* just expressed above. It would imply that anyone who didn't accept the knowledge-norm on belief was guilty of *conceptual incompetence*. That includes a great many people who do seem to be competent with the concept of belief (and many who make the study of belief their life's work). Accusing one's opponents of conceptual incompetence is surely a last refuge. This is not a knock-down argument. I don't claim it is. But it is a significant extra hurdle, and one that defenders of the knowledge-norm explanation might not have expected or wanted to take on.

Exactly the same issues as have been raised for LP3 would also apply to a version with a modal component such as:

**LP4:** It is irrational in the relevant sense to engage in a practice whilst doing something that one believes (or is committed to believing) *necessarily* violates the constitutive norm on that practice.

Both LP3 and LP4 also face an additional challenge. As with LP1 and LP2, neither LP3 nor LP4 are basic or fundamental truths. Both would, if true, require some more fundamental truth to ground them. *Why* are they true (if they are)? What makes them

<sup>18</sup> Shah's truth-norm on belief is formulated in roughly this way: as obtaining by conceptual necessity.

<sup>19</sup> Somewhat ironically in the present context this point has been clearly articulated and championed by Timothy Williamson in his discussion of Van McGee on meaning-constituting views inference-rules. Williamson (2008, pp. 92–94).

true? I leave this as a challenge at this point as I did when discussing LP1 and LP2. I return to it when discussing rational requirements below.

Consider another alternative direction in which we might modify the above linking principles. Focus again on:

**LP2:** It is irrational in the relevant sense to engage in a practice whilst doing something that *necessarily* violates the constitutive norm on that practice.

As with all of the previous linking principles, LP2 begins with ‘It is irrational in the relevant sense...’. This was part of the reason for rejecting LP2: it isn’t obviously irrational in the relevant sense to believe a necessary falsehood. One might, however, try to tone this down. Arguably the following *is* true:

**LP5:** It is *badly defective* to engage in a practice whilst doing something that necessarily violates the constitutive norm on that practice.

Perhaps we can use this as a linking principle. The argument that we would end up with is the following:

1. One cannot know an instance of MP.
2. It is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).
3. It is badly defective to engage in a practice whilst doing something that necessarily violates the constitutive norm on that practice. (LP5).
4. (1–3) It is badly defective to believe an instance of MP.

The conclusion here is weaker than that of previous arguments. It is that believing an instance of MP is badly defective. This proposal has obvious strengths. Most obviously it does not entail, as LP2 does, that it is irrational in the relevant sense to believe necessary falsehoods.

The proposal also faces obvious problems. It fails to differentiate what’s wrong with believing an instance of MP from what’s wrong with believing a necessary falsehood. Both are ‘badly defective’ by this argument. Surely this is a bad result. The sense in which one goes wrong by believing a mathematical falsehood is different from the sense in which one goes wrong by believing an instance of MP. One might respond to this worry by arguing that we also all *believe* or *are committed to believing* that instances of MP violate the constitutive norm (contrasting this with beliefs in necessary falsehoods). That is to say, one might modify LP5 to:

**LP6:** It is badly defective to engage in a practice whilst doing something that *one believes* (or is committed to believing) necessarily violates the constitutive norm on that practice.

But this response runs into the problems already faced by LP3 (and LP4) above. The resulting argument, were one to make this modification, would be:

1. One cannot know an instance of MP.
2. It is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).

3. It is badly defective to engage in a practice whilst doing something that *one believes* (or is committed to believing) necessarily violates the constitutive norm on that practice. (LP6).
4. (1–3) It is badly defective to believe an instance of MP.

The problem—which should be familiar from above—is that this argument is invalid. The conclusion does not follow from 1–3. This is because whilst 2 is about what one ought to do, 3 is about what one believes one ought to do. In addition to this familiar problem, LP5 and LP6 face an additional problem. The problem is really a variant on the challenge raised to LP1–LP4 above to identify the grounds of these linking principles. Just focus on LP5. *In what sense* is it 'badly defective' to violate the constitutive norm on a practice? The obvious answer would be that there is something *rationally* problematic about believing an instance of MP. In order to make this claim we would need to appeal to some principles of rationality. I turn to this below.

#### 4 Rational requirements and the knowledge-norm based explanation

We have run through some of the candidate link-principles to which defenders of the knowledge-norm explanation might appeal. None is fully satisfactory. At the very least each comes with burdens that defenders of the knowledge-norm based explanation may be surprised to take on, and may not want to. We will now look at whether appealing to *requirements of rationality* could help. One might think that they could help either by serving as link principles in their own right, or as supplements to or more basic explanatory grounds for existing candidate principles (such as LP5 and LP6).

I now run through several candidate principles of rationality and suggest that none is of much help. Let's begin by looking at two *bad* candidate rational requirements:

**R1:** One is rationally required to: Follow the constitutive norms of any practice in which one is engaged.

**R2:** One is rationally required to: Believe *p* only if one knows *p*.

Begin with R1. If R1 were true, then—if the knowledge-norm is true—it would be rationally impossible to believe an instance of MP. Believing an instance of MP would violate the constitutive norm on belief and so, by R1, be rationally impermissible. This is an obvious strength of the view. But it faces an equally obvious problem: it is false. Suppose that I am engaged in some practice that is a candidate for having constitutive norms, such as a sport or game. Now suppose I violate one of those norms. Or suppose I violate all of them. I have gone against the rules of the game. But I have not necessarily done anything irrational. This is true on pretty much *any* way of thinking about what rationality is. But it is even clearer if we think of rationality—and rational requirements—much as Broome does: as coherence requirements on mental states. Clearly, in violating the constitutive

norms of a practice, one does not thereby manifest an incoherence in one's mental states.

Much like R1, the strength of R2 is that it would entail that believing an instance of MP is irrational. But R2 is also problematic. Like R1 it is not obviously a good candidate for being a requirement of rationality on a coherence conception. Nor, if it were, would appealing to it be a good move for a defender of the knowledge-norm explanation. There are two reasons for this.

Firstly, consider the conception of rationality that *would* yield R2. R2 is not yielded by a mere 'coherence' conception of rationality à la Broome. A stronger, *substantive* conception of rationality would be required to yield R2. But if we embrace a substantive conception of rationality that yields R2, why would it stop there (i.e. at R2)? Why wouldn't it simply *directly* yield the following as a requirement of rationality: don't believe [p, but I don't believe p]? That is to say, why wouldn't it directly yield the rational impermissibility of believing instances of MP? This requirement and R2 seem pretty much 'on a par' as regards status as candidates for substantive rational requirements. Clearly though, if there were an underived rational requirement not to believe instances of MP, then R2 and indeed the knowledge-norm explanation itself would be redundant. We would already have a simple grounds for explaining what's wrong with believing instances of MP. And that is obviously bad news for a defender of the knowledge-norm explanation.

The second response is even stronger. Suppose that R2 is true (and simply *set aside* the worry that any conception of rationality that entails it would also entail a direct rational requirement against believing instances of MP). The knowledge-norm explanation would now be redundant. It would be redundant because we could simply explain the rational impermissibility of believing an instance of MP in terms of R2. We wouldn't need to appeal to the knowledge-norm on belief at all. The explanation would go as follows:

1. One cannot know an instance of MP.
2. One is rationally required to: Believe p only if one knows p. (R2).
3. (1, 2) One is rationally required not to: Believe an instance of MP.

This explanation has proceeded without any reference to R2. This means that if R2 were true, then the knowledge-norm explanation would be out-of-business.<sup>20</sup>

Hopefully seeing why R1–R2 are bad candidates is useful. Let's now look at what one might think to be a *better* candidate: an enkratic requirement. Roughly:

**R3:** One is rationally required to: If one believes one ought to believe p, then believe p.<sup>21</sup>

<sup>20</sup> One could argue that if R2 were true, then something very much like the knowledge-norm would actually be true. However R2 is a principle of rationality that does not refer to, presuppose or entail a principle that, like the knowledge-norm, makes a claim about what is constitutive of belief.

<sup>21</sup> I do not consider all alternative formulations of the enkratic requirement here. One may, for example, reformulate it in negative terms as follows. One is rationally required to: If one believes one ought not to believe p, then not believe p.

Unlike R1–R2, R3 is a widely (though not universally) accepted principle or rationality. I will simply grant this. How could it help to show that believing an instance of MP is irrational? One might think that it could do so by showing that believed violations of the knowledge-norm are *rational* violations. The basic argument would be as follows:

1. One cannot know an instance of MP.
2. It is constitutive of believing a proposition that one ought to do so only if one knows it (*knowledge-norm*).
3. One is rationally required to: If one believes one ought to believe p, then believe p (*enkrasia*).
4. (1–3) One is rationally required not to: believe an instance of MP.

This would be a good conclusion for a defender of the knowledge-norm-based explanation. It would show that believing an instance of MP is necessarily akratic. But the argument is invalid. It is invalid for a reason familiar from the discussion of LP3, LP4 and LP6 above. The second premise is about what one ought to believe. The third premise—*enkrasia*—is about what one *believes* one ought to believe. This is an extra commitment that goes well beyond the knowledge-norm: it is neither entailed by the knowledge-norm, nor does it entail the knowledge-norm, nor is it obviously true.

A further point about this is in order. Suppose that one *does* appeal to an enkratic principle according to which one is rationally required to: if one believes one ought to believe p, then believe p. There are two ways of reading the 'ought' in this requirement. To see this we must return to the distinction made at the beginning of the article between *genuinely normative* oughts or requirements and *mere correctness norms*.

The 'ought' that figures in enkratic principles is a *genuinely normative* ought.<sup>22</sup> It cannot be a mere correctness norm. This is, I take it, uncontroversial. It is not the case that I am rationally required to: if I judge that it is correct according to the standards of chess for me to make a particular move, then make that move. There may nothing at all to be said—normatively—for adhering to the norms of chess. And so there may be no normative reason for me make the correct move. If this is the case, and if I judge it to be the case, then I am not rationally required to make the move (even if we read the requirement as wide-scope).

This means that the 'ought' in 2\* must be interpreted as a genuinely normative ought rather than a mere correctness norm. This isn't a big problem in itself: it is of course possible to interpret it in this way. It just puts an extra burden on someone who would appeal to *enkrasia* as the basis of the knowledge-norm explanation. It is an extra burden because constitutive norms on belief—whether truth-based or knowledge-based—are sometimes formulated quite explicitly as mere correctness norms.<sup>23</sup> And this is not without justification. After all, the constitutive norms of

<sup>22</sup> See e.g. Broome (2013).

<sup>23</sup> See (e.g.) Cote-Bouchard (2016) and Cote-Bouchard and Littlejohn (2018). It is also possible to read Steglich-Petersen (2011) in this way.

ordinary practices—such as sports and games—do seem to be the kinds of thing that are best understood in terms of mere correctness. Generally, the point is that constitutive norms are the kinds of things that have ‘normative force’ (i.e. that one has normative reason to follow) only if one has some normative reason to engage in the practice of which they are constitutive in the first place. A defender of the knowledge-norm explanation that is based on *enkrasia* must claim that the ‘ought’ in 2\* is not like this. She must claim that it is genuinely normative. I won’t enter into the complicated arguments regarding this reading here. It would take us too far afield.<sup>24</sup> I simply note that it is an extra commitment.

## 5 Conclusion

I have tried to articulate how norm-based explanations of Moore’s Paradox are meant to work and to outline a problem that they face. The problem comes from the need to appeal to linking-principles: principles that take us from violations of constitutive norms on belief to explanations of Moore’s Paradox. I have tried to find a plausible linking-principle. I have not succeeded. Each comes with significant costs. The principles LP1–LP6 fail in either of two ways. Firstly, they fail to properly differentiate the sense in which it is problematic to violate MP from the sense in which it is problematic to believe a falsehood or a necessary falsehood (LP1, LP2, LP5). Secondly, they rely not on the knowledge-norm itself but on a belief in that norm—something that neither entails nor is entailed by the norm (LP3, LP4, LP6). In addition to these two causes of failure, none of the principles LP1–LP6 possesses a sufficient degree of independent motivation: they would need some more basic principle or principles to ground them. One might be optimistic that R1–R3 could promise to provide this. But these principles also fail in two ways. Firstly, R1 and R2 are too strong to be independently plausible. And anyhow if they were true they would arguably render the knowledge-norm explanatorily redundant. Secondly, R3, like LP3, LP4 and LP6, relies not on the knowledge-norm itself but on a belief in that norm. It also relies on an interpretation of the knowledge-norm as stronger than mere correctness (R3).

I do not conclude from this that norm-based explanations could not work or that there couldn’t be a satisfactory link-principle. I do not have a general argument for such a strong conclusion. But I do hope to have drawn attention to the problem that norm-based explanations face; a problem that has been largely missed in the literature.<sup>25</sup> Solving it is an interesting challenge for defenders of the knowledge-norm explanation of Moore’s Paradox. What are their options? The first option is to significantly weaken the characterisation of what is wrong with believing an instance of MP, rendering it equivalent to belief in a necessary falsehood. This strikes me as deeply unsatisfactory. The second option is to argue that the

<sup>24</sup> One of the substantial relevant issues here is whether believing is the kind of practice for which one necessarily possesses reasons for engagement. For some sceptical discussion see Enoch (2011).

<sup>25</sup> For an attempt to make the point generalise see Greenberg and Cowie (2018).

knowledge norm is not only true, but is also believed to be true. This is more palatable but would still be a difficult sell. It would also need to be supplemented with both a strong interpretation of the normative relation referred to in the knowledge-norm and a strong interpretation of the enkratic principle (an issue I have not touched on here). The third option is to come up with some radically different alternative. I leave this as a challenge to defenders of the knowledge-norm.

**Acknowledgements** I am grateful to Alexander Greenberg for discussion. This paper develops themes from Greenberg and Cowie (2018). I am especially grateful to audiences at the Universities of Glasgow and Oxford for feedback and to an anonymous reviewer for *Philosophical Studies* for helpful comments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Baldwin, T. (2007). The normative character of belief. In M. S. Green & J. N. Williams (Eds.), *Moore's paradox: New essays on belief, rationality, and the first person*. Oxford: Clarendon Press.
- Boghossian, P. (2003). The normativity of content. *Philosophical Issues*, 13(1), 31–45.
- Broome, J. (2007). Is rationality normative? *Disputatio*, 2(23), 161–178.
- Broome, J. (2013). *Rationality through reasoning*. Hoboken: Wiley-Blackwell.
- Bykvist, K., & Hattiangadi, A. (2007). Does thought imply ought? *Analysis*, 67(296), 277–285.
- Carter, J. A., Gordon, E. C., & Jarvis, B. (2017). *Knowledge first: Approaches in epistemology and mind*. Oxford: Oxford University Press.
- Cote-Bouchard, C. (2016). Can the aim of belief ground epistemic normativity? *Philosophical Studies*, 173(12), 3181–3198.
- Cote-Bouchard, C., & Littlejohn, C. (2018). Knowledge, reasons and errors about the error theory. In R. McKenna & C. Kyriacou (Eds.), *Metaepistemology: Realism and antirealism*. Basingstoke: Palgrave MacMillan.
- de Almeida, C. (2001). What is Moore's paradox about? *Philosophy and Phenomenological Research*, 62(1), 33–58.
- Douven, I. (2006). Assertion, knowledge, and rational credibility. *The Philosophical Review*, 115(4), 449–485.
- Douven, I. (2009). Assertion, Moore, and Bayes. *Philosophical Studies*, 144, 361–375.
- Engel, P. (2014). In defense of normativism about the aim of belief. In T. Chan (Ed.), *The aim of belief*. Oxford: Oxford University Press.
- Enoch, D. (2011). Schmagency revisited. In M. Brady (Ed.), *New waves in metaethics*. Basingstoke: Palgrave MacMillan.
- Green, M., & Williams, J. (2007). Introduction. In M. Green & J. Williams (Eds.), *Moore's paradox*. Oxford: Oxford University Press.
- Greenberg, A., & Cowie, C. (2018). Constitutivism about epistemic normativity. In C. Kyriacou & R. McKenna (Eds.), *Metaepistemology: Realism and anti-realism*. Basingstoke: Palgrave MacMillan.
- Howard, C. Forthcoming. The fundamentality of fit. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics*.
- Huemer, M. (2007). Moore's paradox and the norm of belief. In S. Nuccetelli & G. Seay (Eds.), *Themes from G.E. Moore: New essays on epistemology and ethics*. Oxford: Clarendon Press.
- Littlejohn, C. (2010). Moore's paradox and epistemic norms. *Australasian Journal of Philosophy*, 88, 79–100.
- Lo Presti, P. (2014). Moore's paradox and epistemic norms. *Logos & Episteme*, 5(4), 445–464.

- McHugh, C. (2012). The truth norm of belief. *Pacific Philosophical Quarterly*, 93(1), 8–30.
- McHugh, C., & Way, J. (2016). Fittingness first. *Ethics*, 126(3), 575–606.
- Olson, J. (2014). *Moral error theory: History, critique, defence*. Oxford: Oxford University Press.
- Shah, N. (2003). How truth governs belief. *The Philosophical Review*, 112(4), 447–482.
- Shah, N., & Velleman, D. (2005). Doxastic deliberation. *The Philosophical Review*, 114(4), 497–534.
- Shoemaker, S. (1996). Moore's paradox and self-knowledge. *Philosophical Studies*, 77, 211–228.
- Smithies, D. (2012). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research*, 85(2), 273–300.
- Sorensen, R. (2000). Moore's problem with iterated belief. *Philosophical Quarterly*, 50, 28–43.
- Steglich-Petersen, A. (2011). How to be a teleologist about epistemic reasons. In A. Steglich-Petersen & A. Reisner (Eds.), *Reasons for belief*. Cambridge: Cambridge University Press.
- Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 16, 267–297.
- Wedgwood, R. (2007). *The nature of normativity*. Oxford: Oxford University Press.
- Whiting, D. (2010). Should I believe the truth? *Dialectica*, 61, 213–224.
- Williams, J. (2013). Moore's paradox and the priority of belief thesis. *Philosophical Studies*, 165, 1117–11138.
- Williams, J. (2015a). Moore's paradox in thought: A critical survey. *Philosophy Compass*. <https://doi.org/10.1111/phc3.12188>.
- Williams, J. (2015b). Moore's paradox in speech: A critical survey. *Philosophy Compass*. <https://doi.org/10.1111/phc3.12187>.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Williamson, T. (2008). *The philosophy of philosophy*. Oxford: Oxford University Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.