Christopher Cowie
Alexander Greenberg

Constitutivism about Epistemic Normativity

For *Metaepistemology: Realism and Antirealism,* Palgrave MacMillan. Forthcoming, 2018. Please cite final version.

Suppose that you possess strong, perhaps decisive, evidence for a proposition that you're considering. It is natural to think that that you thereby possess a *reason* to believe that proposition. What explains this? Why is there such a connection between evidence for (or against) propositions and reasons for (dis)believing them? What, in other words, are the *grounds* of epistemic normativity?

There are a number of different candidate answers to this question. One candidate is that epistemic normativity doesn't have any grounds. According to this view, there are just some basic – that is ungrounded – truths about what we ought to believe. At the other end of the spectrum is the view that epistemic normativity is grounded in practical normativity. According to this view, epistemic reasons for belief are explained by the contribution that believing in light of the evidence makes to prudential, moral or aesthetic ends. Neither of these views is ideal. The former isn't much of an *explanation* at all. It makes epistemic normativity a fundamental constituent of reality.[1] The latter struggles with the fact that believing evidentially supported propositions isn't always of practical value, and the fact that it sometimes seems to be of practical value to have evidentially unsupported beliefs. Of course, defenders of these views have developed sophisticated responses to these basic charges. But the fundamental problems remain. It would be ideal if there were some happy middle-ground; if epistemic normativity were neither ungrounded, nor grounded in the practical. That could allow us to avoid the problems that each approach faces. Interestingly, there is a candidate for occupying this middle ground. The candidate is that epistemic normativity is grounded in the nature of the attitude of *belief*. This view is now often referred to as *constitutivism*, or, more fully, *constitutivism about epistemic normativity*. The idea is that *the nature of belief* explains why have reasons to believe, or disbelieve, propositions in light of evidence possessed for, or against, them.

How is this meant to work? The idea is that *the concept of belief is normative*. This means that for an attitude to fall under the concept 'belief' is just for it to stand in normative relations to a certain property (or properties). This property is said to be the *correctness condition* for belief. Belief's correctness condition is typically claimed to be truth; specifically, the truth of the proposition believed. We can therefore express constitutivism about epistemic normativity as follows:

> **TRUTH NORM**: It is a conceptual truth about belief that it is correct to believe a proposition if and only if it is true.[2]

---

[1] Though for a defence of something close to this view, see Kyriacou (this volume).

[2] It should be noted that one could claim that a belief is correct just in case it's true without being committed to TRUTH NORM. TRUTH NORM also claims this norm is *constitutive* of belief. We are concentrating on this latter claim, because only it would be able to provide the grounds of epistemic normativity. If belief is subject to a truth norm but not essentially

Constitutivism offers an explanation *in terms of the nature of belief* for why there are normative relations between belief and truth. And once this in place, it isn't a big leap – the thought is – to explain why there are normative relations between belief and *evidence* (evidence being an indicator of truth). We now have an account of the grounds of epistemic normativity.

Does this constitutivist explanation of epistemic normativity work? Our focus will not be on whether TRUTH NORM, if accepted, can ground epistemic normativity. The aim of this article is to assess the arguments that have been given for TRUTH NORM. We will ask whether these arguments are persuasive, and we will claim that they are not. We will claim that there is insufficient reason to accept TRUTH NORM. *A fortiori* there is insufficient reason to think that TRUTH NORM is the correct explanation of the grounds of epistemic normativity.

There are a range of different arguments for TRUTH NORM (for a survey, see McHugh and Whiting 2014). As such, one might expect that we would have to proceed in a piecemeal fashion, taking each argument for TRUTH NORM in turn. Our ambition, however, is to take a more unified approach. We aim to present a common strategy. We aim to show that some key arguments for TRUTH NORM fail for a common reason. They fail because TRUTH NORM fails to provide a good explanation of the phenomena that it is meant to explain. More precisely, we will claim arguments for TRUTH NORM falter on a dilemma. The dilemma is as follows:

> **DILEMMA**: TRUTH NORM is appealed to as the best explanation of some phenomenon (the explanandum).
>
> (i)   *Either* TRUTH NORM doesn't look like it best explains the explanandum in question.
>
> (ii)   *Or* TRUTH NORM can explain the explanandum, but *only if* it is supplemented by an implausible claim about our motivations to follow prescriptions.

Either way, TRUTH NORM fails to provide the best explanation for the phenomenon in question. Given that the ability to provide such an explanation is the reason for positing TRUTH NORM, we shouldn't posit TRUTH NORM.

This is our DILEMMA. We will apply it in two specific arguments for TRUTH NORM: first to an argument based on *deliberative transparency*; second to an argument for based on *Moore-paradoxical sentences*. Part of what is interesting about our strategy is that we are drawing together these seemingly distinct arguments for and against TRUTH NORM in a simple, unifying way. We will begin (section 1) by making some preliminary remarks about how our approach differs from and, we think, improves upon existing opposition to TRUTH NORM.

---

so, that would be a substantial first-order epistemological fact, but it wouldn't be able to ground epistemic normativity. This point is not often clearly made in the literature on truth norms (though see McHugh 2013, 463; Steglich-Petersen 2013, 279, fn. 2; Greenberg 2017, chap. 2).

We will then (section 2) show how our strategy applies against arguments for TRUTH NORM based on deliberative transparency. We will then (section 3) show how this strategy applies against arguments for TRUTH NORM based on Moore-paradoxical sentences. We then conclude (section 4).

1. What distinguishes our argument from other arguments against TRUTH NORM

The argument that we will develop against TRUTH NORM differs from some of the existing arguments in the literature. In particular there are two important kinds of criticism that have been made of TRUTH NORM that our argument should be distinguished from.

One influential criticism of TRUTH NORM concerns the precise content of the norm that is alleged to be constitutive of belief. There has been considerable debate about whether a truth norm can be formulated in such a way as to avoid implausible prescriptions – such as that one ought to believe ungraspably complex truths and a prescription believe 'true blindspots', (Bykvist and Hattiangadi 2007, 2013). We will set this aside however, and grant that the defender of TRUTH NORM has some response to this criticism.

A second important criticism of TRUTH NORM is the 'correctness' of beliefs it refers to is not 'heavy duty' enough to do any epistemological or metaphysical work (Côté-Bouchard 2016; cf. Rosen 2001, 619–21; Bykvist and Hattiangadi 2013, sec. 2). How our argument differs from this is somewhat complex. While our conclusion is the same – that TRUTH NORM does not best explain the nature of belief – our argument targets a later stage in the dialectic. The arguments we will be rejecting – which claim that TRUTH NORM is required to explain deliberative transparency and Moore's paradox – can be understood as attempts to argue, *contra* the above criticism, that belief's correctness conditions have to be understood as *genuinely normative*. Our conclusion is that this is not true: these phenomena do not mandate a genuinely normative understanding of belief's correctness conditions.

2. Does TRUTH NORM explain deliberative transparency?

One influential argument for TRUTH NORM is that it best explains the transparency of doxatic deliberation. We will begin by outlining this argument, largely following the presentations of it given by Nishi Shah and David Velleman (Shah 2003, 260–74; Shah and Velleman 2005, 500–502), before criticising it through our DILEMMA (in §§2.1-2.2).

When we deliberate as to whether to believe that p, we recognise that this question inevitably gives way to the question of whether p is true. Furthermore, not only is the deliberative question of whether to believe that p answered by the factual question of whether p is true, it is *only* answered by that question, to the exclusion of any other considerations, such as whether believing that p would be in one's interests. This is what it means to say that deliberation about what to believe is *transparent*. What explains this phenomenon? Arguably the obvious place to look for an explanation is to the existence of a tight relationship between belief and truth, because the phenomenon just is one of a question about *belief* giving way to a question about truth. TRUTH NORM prescribes that we believe p if and only if p is true. Consequently, when we deliberate about whether to believe p, we will – if we follow that norm – simply deliberate about whether p is true. The

transparency of deliberation is explained. So TRUTH NORM can explain the transparency of deliberation.

This shows that TRUTH NORM is a sensible way of explaining transparency. But why think it is the *best* explanation? The main reason given by its defenders is that it provides a better explanation than alternative *dispositional* accounts of belief. To explain the *exclusive* focus on truth in deliberation, the dispositions specific to belief (that dispositional accounts posit) would need to dispositions to form, maintain, and revise beliefs in and, crucially, *only in* response to truth-relevant considerations. Call this the *strong dispositional account*.

The fact that a dispositional account of belief would have to be 'strong' in this respect in order to explain transparency is the reason why TRUTH NORM is claimed to provide a better explanation. Take a step back. Note that although when we *deliberate* about whether to believe p we simply consider whether p is true, not *all* of our beliefs are like this. Sometimes, what we believe is determined by our desires and other conative states, in cases of wishful thinking. The key datum, according to defenders of TRUTH NORM, is that wishful thinking doesn't exert its influence *within conscious deliberation*.

This presents a problem for the strong dispositional account. On this account, the products of wishful thinking would not count as beliefs. Now one could respond to this by *weakening* the dispositional connection between belief and truth so as to allow for the products of wishful thinking to count as beliefs, and defend what we can call the *weak dispositional account*. But a defender of this view can no longer explain why, when we consciously deliberate about whether to believe that p, we simply deliberate as to whether p is true. On her account, it should, in principle, be possible to deliberate about whether p, and be moved in those deliberations in part by non-truth related (e.g. conative) considerations. But, *ex hypothesi*, it's not. So this weak dispositional account fails too.

In fact, we can represent the failings of both of the dispositional accounts as follows:

	Explains Transparency in Deliberation?	Allows for the Possibility of (Non-Del.) Wishful Beliefs?
Strong Dispositional Account	Yes	No
Weak Dispositional Account	No	Yes

Neither of these accounts works because neither can answer 'yes' to both questions. This is where the real strength of TRUTH NORM comes in. TRUTH NORM seems able to answer 'yes' to both:

TRUTH NORM	Yes	Yes

Why is this? The key point is that TRUTH NORM is a view about the *concept* of belief. This is key because and when we deliberate about what to believe, the concept of belief is playing a role in framing what we do, whereas it isn't in ordinary non-deliberative belief-formation. It is this that allows TRUTH NORM to explain transparency in deliberation, whilst still allowing for the possibility of wishful thought. And so, it is this that makes TRUTH NORM the *best* explanation of transparency in deliberation.

2.1 Transparency in deliberation: Our DILEMMA

We don't think that this is a good argument for TRUTH NORM. We can show why by appealing to the following DILEMMA:

> **DILEMMA**: TRUTH NORM is appealed to as the best explanation of the transparency of deliberation. But (we argue):

> (i) *Either* TRUTH NORM doesn't look like it best explains the explanandum in question.

> (ii) *Or* TRUTH NORM can explain the explanandum, but *only if* it is supplemented by an implausible claim about our motivations to follow prescriptions.

We will make the case quickly before expanding on it in what follows. The suggestion that this argument for TRUTH NORM has a problem with *motivation* was first suggested by Asbjørn Steglich-Petersen (2006), and was further developed by Conor McHugh (2013). But we extend this point, and, crucially, in the following section we attempt to show that is generalises to attempts to use TRUTH NORM to respond to Moore's paradox.

Let's begin with horn (ii). According to defenders of TRUTH NORM, deliberative transparency works as follows. The concept of belief prescribes believing p iff p is true. Deliberation about what to believe is framed by the concept of belief, according to defenders of TRUTH NORM, because it aims to an answer the question *whether to believe that p*. So, given that it's part of the concept of belief that a belief is correct iff it is true, a deliberator about what to believe will necessarily take themselves to be under this prescription. So they will try to accept the proposition about which they are deliberating iff it is true. This explains why their deliberation is transparent; it explains why the only consideration they regard as relevant is p's truth, and not, for example, whether believing p is in their interest.

One of the obvious problems with this story is that it seems to require a very strong connection between (a) taking oneself to be under a prescription to behave in some way, and (b) actually behaving in that way. It seems to require that because deliberators will take themselves to be under a prescription to believe that p iff p is true, they will actually try to do that. We can label it as follows:

> (1) If you take it to be correct for you to V, then you will be motivated to V.

This is a form of 'judgement internalism'. But it isn't a plausible form. Frequently, we take ourselves to be under a prescription to behave in some way, yet we don't try to behave in that way. So (1) looks like a bad principle. It's far too strong to be true. This is borne out by considering the forms of judgement internalism actually defended by moral philosophers. Those who claim that there is some necessary connection between judging an action to be morally right (or good, correct, or what one ought to do) and being motivated to perform that action typically claim that this connection is, in various ways *defeasible*. Some claim that the moral judgement necessarily results in *some* motivation to act, but not necessarily

overriding motivation; others claim that moral judgement necessarily results in motivation to act in *ideal* circumstances, or if one is *rational* (for a survey, see Svavarsdottir 1999, sec. 1). A representative example of this kind of judgement internalism is put forward by Michael Smith:

> "[A]gents who judge it right to act in various ways are so motivated, and necessarily so, absent the distorting influences of weakness of the will and other similar forms of practical unreason on their motivation." (Smith 1994, 61)

These qualified versions of judgement internalism are typically defended, as the passage from Smith indicates, because a version of internalism without such qualifications would have the implausible consequence that it would be impossible to judge that something is right but fail to do so out of weakness of the will, deep depression, etc. But these kinds of cases of irrationality also show that a principle like (1) – which expresses a simple entailment between judging it to be correct to V and being motivated to V – is implausible.

Can the defender of TRUTH NORM accept that form of internalism expressed in (1) is implausibly strong? This leads us onto horn (i). Perhaps, she might argue, it is possible that we do sometimes take ourselves to be under the prescription of TRUTH NORM, yet fail to follow it. The result is that in some cases we will be able to deliberate as to whether to believe that p, yet not be sensitive only to truth-directed related considerations. But, the defender of TRUTH NORM might claim this is ok because deliberation isn't always transparent. This would be true if there are certain cases of self-conscious *epistemic akrasia*, in which someone judges that the evidence suggests that p is true, but doesn't believe that p. In fact, Pascal Engel defends the claim that the truth norm need not necessarily motivate by appealing to cases of this kind:

> "Just as cases of *akrasia* or *accedia* can arise where the agent considers the norm but does not follow it, cases where the norm of truth is considered by the agent but is not followed can arise." (Engel 2007, 199, cf. 2006, 369–72)

If this is correct, then the defence of TRUTH NORM need no longer rely on a very strong connection between the prescriptions that we take ourselves to be under, and our tendency to follow them. This is because the explanandum has now changed; what is now to be explained is a general tendency for deliberation to be transparent. But the problem is that TRUTH NORM now loses its unique explanatory power. A weak dispositional account – one which claims that belief is essentially responsive to be truth-relevant considerations, but not exclusively so – can explain a mere general tendency for deliberation to be transparent. We don't need TRUTH NORM any more.[3]

That's the general structure of the DILEMMA. In horn (ii) a defender of TRUTH NORM is shown to rely on a *very* strong connection between putting oneself under a prescription to behave in some way, and behaving in that way. And in trying to escape horn (ii) the explanandum (transparency in deliberation) has to be weakened to avoid this commitment.

---

[3] This point is made by Andrei Buleandra (2009, 328–31).

This then lands the defender of truth norm on horn (i) because it results in TRUTH NORM losing its unique explanatory appeal.

What are the best lines of response to our DILEMMA? We will argue that all the obvious moves that might be made by the defender of TRUTH NORM still fall on one of the two horns. Either they involve implausibly strong claims about motivation, or they have the result that TRUTH NORM does not best explain transparency.

2.2 Can the defender of TRUTH NORM claim epistemic akrasia is impossible?

One move the defender of TRUTH NORM could make is to claim that the strong version of internalism expressed by (1) is not, as we claim, implausible if it is just restricted to belief. One might argue, according to this line of thought, that (1) is implausible in relation to *action*, because practical akrasia – self-consciously acting contrary to what one judges to be correct to do – is a real and familiar phenomenon. However, one could claim that (1) is plausible *if* restricted to belief. One might try and motivate this claim by arguing that epistemic akrasia – self-consciously believing contrary to what one judges to be correct – is not in fact possible. This kind of claim is made by Philip Pettit and Michael Smith:

> "Imagine that your beliefs run counter to what evidence and fact require. In such a case, your beliefs will not allow those requirements to remain visible because the offending beliefs themselves give you your sense of what is and your sense of what appears to be. You are therefore denied an experience whose content is that you are believing such and such in defiance of the requirements of fact and evidence. (Pettit and Smith 1996, 448)

Pettit and Smith claim that epistemic akrasia is impossible because, roughly, one's judgement about the correctness of one's belief that p just is one's belief about whether p. This provides a response to our DILEMMA is because this alleged impossibility of epistemic akrasia could be claimed to support the strong form of internalism expressed in (1) (if restricted to belief). This is because it would show that judging that it's correct to believe that p does in fact entail being motivated to believe that p.[4]

We don't think this escapes our DILEMMA. Our response is that the claim that epistemic akrasia is impossible is an equally contentious and implausibly strong claim about motivation as (1). As such, it still falls foul of horn (i).

Why is Pettit and Smith's claim implausible? While it correct that one's judging that one's belief that p is correct or incorrect *typically* results in one forming or revising a belief that p, it's quite another thing to claim these are *just one and the same*. This second claim is what

---

[4] To our knowledge, no one in the literature on deliberative transparency defends the claim that epistemic akrasia is impossible; the closest is Sophie Archer, who defends the claim that it is impossible to both judge one has sufficient evidence for p and refrain from believing p (2017). Some also doubt whether judging one's own beliefs to be incorrect suffices for epistemic akrasia (Owens 2002), but this does not matter for our purposes. Our concern is whether such a state is possible, not whether it is akrasia accurately so-called.

denying the possibility of epistemic akrasia has to involve; but this second claim is implausibly strong. It is implausibly strong because it rules out certain cases – like Hume's judgement that even though he lacks reason for his beliefs in the external world, he can't help having them – which may be *irrational* but certainly aren't impossible. In this way, the option of denying the possibility of epistemic akrasia lands the defender of TRUTH NORM on horn (i) of our DILEMMA, because it relies on a claim about motivation that is just as implausibly strong as the internalism expressed in (1).

2.3 Can the defender of TRUTH NORM appeal to a more plausible form of internalism?

The other obvious move for the defender of TRUTH NORM is to claim that their explanation of transparency only needs to appeal to a weaker – and therefore more plausible – version of internalism. But we will show that weaker forms of internalism are still vulnerable to our dilemma; what they gain in plausibility, they lose in their ability to explain transparency.

The clearest way to amend the implausibly strong form of internalism expressed in (1) is as follows:

(2) If you're rational, then: if you take it to be correct for you to V, then you will be motivated to V.

This is a much weaker – and therefore more defensible – account of the relation between taking oneself to be under a prescription and following it.

However, if they are going to accept (2), then defenders of TRUTH NORM will also need to change how they characterise what TRUTH NORM is supposed to explain. Because (2) is restricted to *rational* motivation, it will only be able to explain the transparency of *rational* doxastic deliberation. In other words, the explanandum now becomes the fact that *if you're rational*, then if you deliberate about whether to believe that p, your deliberation will be transparent. This could allow defenders of TRUTH NORM to explain deliberative transparency with the weaker version of judgement internalism expressed in (2).

Can moving to (2) save TRUTH NORM from the above DILEMMA? We don't think it can. This is where horn (i) of our DILEMMA bites. The more defensible version of internalism expressed in (2) means that explanandum needs to be recharacterised as the transparency of *rational* doxastic deliberation. But recharacterising the explanandum in this manner means that TRUTH NORM loses its unique explanatory power, because other accounts can equally well explain the transparency of *rational* doxatic deliberation.

Specifically, recall the strong dispositional account, which claimed that belief is essentially regulated by dispositions to be responsive to be formed, maintained, and revised in *and only in* response to truth-relevant considerations. Because the explanandum has now been changed to the fact that *rational* doxastic deliberation is transparent, in order to explain that these dispositions need not be claimed to be essential to belief as such; they only need to be claimed to be essential to *rational* belief. And if such dispositions are essential to rational belief, that can perfectly well explain why rational deliberation about what to believe is motivated by and only by truth-relevant considerations. Furthermore, whereas it

was thought to be problematic to think of these strong truth-responsive dispositions as essential to belief *as such* – because it had the result that the products of wishful thinking don't count as belief – it is *not* problematic to think of these strong truth-responsive dispositions as essential to *rational* belief, because the products of wishful thinking aren't *rational* beliefs. In this way, if the explanandum is weakened to the transparency of *rational* doxastic deliberation, an alternative dispositional account can explain the weakened explanandum just as well as TRUTH NORM. Weakening the explanandum in this fashion has the result, therefore, that it is not best explained by TRUTH NORM.[5]

A second weaker form of internalism is in fact offered by Shah and Velleman. They themselves are wary of the strong view of internalism that their view seems to require. So they offer a weaker alternative in a footnote (2005, fn.40, 533-34). But the footnoted version of internalism is also vulnerable to our DILEMMA. On the first horn, their weaker alternative internalism is in fact too weak to explain transparency. On the second horn, they could get around this by weakening the conception of transparency that they are trying to explain. But if they do this then TRUTH NORM isn't a better explanation than competitors.

Let us explain why. Shah and Velleman claim that their argument in fact only relies on a weaker form of internalism than (1), which they outline as follows:

> "The relevant form of internalism does not require a positive disposition to obey any norm that one applies; what it requires is the lack of a disposition to obey a different norm instead. One cannot genuinely apply the norm of truth to an attitude while simultaneously trying *only* to make it conform to some other, unrelated norm. This form of internalism does not rule out obedience to additional norms compatible with the one applied. One can aim to arrive as quickly as possible at a true cognition with respect to p – in which case, one will deliberate in accordance with a norm of speed as well as the norm of truth." (Shah and Velleman 2005, 533–34, italics ours).

We can formulate Shah and Velleman's weaker internalism as follows:

(3) If you take it to be correct for you to V, then you will not be motivated by norms prescribing refraining from V-ing.

(3) is still vulnerable to our dilemma. This is because, according to (3), someone's accepting TRUTH NORM does *not* entail that they are motivated to believe the truth and – crucially – *only* the truth. All that acceptance of TRUTH NORM means, according this version of internalism, is that one cannot be motivated *only* by non-truth-related considerations. But this is consistent with the possibility that one *can* – even if one accepts TRUTH NORM – be motivated *in part* by non-truth-related considerations (cf. McHugh 2013, 457).

---

[5] McHugh makes a similar point when he argues that a weaker constraint – which he calls 'efficacy' – is only a feature of rational doxastic deliberation (2013, 458–62). Our point is more conditional: *if* the internalism expressed in (2) is appealed to, then only the transparency of *rational* doxastic deliberation can be explained by TRUTH NORM, which means it loses *its* unique explanatory power.

But if this is the case, this seems to be too weak to explain transparency. Why? Because transparency seems to be a phenomenon in which you are motivated *only* by truth-relevant considerations.

Now Shah and Velleman could get around this by denying this conception of transparency. They could say that in ordinary or typical doxastic deliberation, we are only motivated in part by truth-relevant considerations. But if they say this, then TRUTH NORM isn't a better explanation than competitors, in particular, the weak dispositionalist account, and they succumb to horn (i). So Shah and Velleman's own weakened internalism is also vulnerable to our DILEMMA.

Both of these weaker forms of internalism, therefore, suffer from our dilemma. They *either* mean that TRUTH NORM cannot explain transparency, *or* if transparency is recharacterised so that TRUTH NORM can explain it, it no longer *best* explains it. There are, or course, alternative weaker versions of internalism we haven't discussed. One might claim that judging it is correct for one to V entails being motivated to V *somewhat* or *in ideal circumstance*. But there is every reason to think that other weaker versions will be just as vulnerable to horn (i) of our DILEMMA, for predictable reasons. That is, they will require us to weaken the explanandum in such a way that TRUTH NORM no longer uniquely explains it.

3. Does TRUTH NORM explain the absurdity of accepting Moore Paradoxical Sentences?

In the preceding sections we have explained how our DILEMMA works against the claim that TRUTH NORM is the best explanation of deliberative transparency. We will now extend our basic argumentative strategy in a novel direction. We will argue that our DILEMMA can be used to respond to attempts to use Moore-paradoxical sentences to argue for TRUTH NORM. The basic structure of our argument mirrors those used above. Attempts to use TRUTH NORM to explain Moore-paradoxical sentences are vulnerable to our DILEMMA. Either they are inadequate explanations of the explanandum, or they must be supplemented by implausibly strong claims about our motivations to follow prescriptions.

Moore-paradoxical sentences are sentences of the following form.

  MP: p, but I don't believe that p.

These sentences can sometimes be true. Suppose that Donald Trump were to say: [man-made climate change is happening, but I don't believe that man-made climate change is happening]. This would be true. It would be true because both of its conjuncts are true. The first conjunct is true: man-made climate change is happening. And the second conjunct is true: Trump doesn't believe that it is happening (at least according to his Twitter account).

Although instances of MP can be true, it is often thought that there would be something absurd, incoherent or even impossible with one's accepting – whether believing or sincerely asserting – an instance of it from one's own, first-personal perspective. In sincerely saying, or thinking: [man-made climate change is happening, but I don't believe that man-made

climate change is happening] Trump would be doing something absurd, incoherent, or perhaps even impossible.

In the philosophical literature on these sentences, the absurdity – in some form or other – is taken for granted. The philosophical trick is to explain its nature and source. The explanation, we know, can't appeal to the inconsistency of the two conjuncts. We know that they're not necessarily inconsistent. So how can we explain it?

One popular kind of explanation appeals to features of *speech*. It is absurd to say, or assert, instances of MP because the first conjunct of MP conversationally implies the falsity of the second or is inconsistent with some conversational principle or maxim, or some principle or maxim governing the nature of assertion itself. Whilst there is a lot to be said for this kind of approach, it faces a problem. The problem is that, prima facie, the absurdity of accepting a Moore-paradoxical sentence isn't restricted to *saying* or asserting it. It's also absurd to genuinely *think* it. Any explanation of the absurdity of accepting Moore-paradoxical sentences needs to respect this. The point is well-put by Michael Huemer:

> "Moore's Paradox cannot be fully resolved by appeal to rules governing solely the use of language, because it is easy to construct non-linguistic versions of the paradox. It would be absurd to think to oneself… the thought, that it is raining but that one does not believe this." (Huemer 2007, 144)

Bearing this in mind, it isn't surprising that explanations of the absurdity of accepting instances of MP should draw on the nature of belief. With that in mind, it is interesting to think about TRUTH NORM. Can TRUTH NORM help to explain the absurdity of accepting an instance of MP? You might think (or at least suspect) so. Certainly, there are lots of references in the literature to the relation between the normativity of belief and the absurdity of Moore-paradoxical sentences. For example, both Bernard Williams and Peter Railton claim that Moorean absurdity is a product of the fact that belief 'aims at truth' (Williams 1970, 137; Railton 1994, 72–73), and those who defend TRUTH NORM often claim it is a way of unpacking the idea that belief aims at truth (see, e.g., Wedgwood 2002, 267; Engel 2013, 32). It's not hard to get a rough sense that this might be right.

The closest to a developed argument for TRUTH NORM on the grounds that it explains Moore's paradox is given by Thomas Baldwin (2007). Baldwin, however, phrases his explanation in terms of believing that p essentially involving a 'normative commitment to truth' rather than in terms of a truth norm. On Baldwin's account, Moorean absurdity is "the absurdity of embracing, within a single assertion or judgment, a commitment both to the truth of *p* and to one's being committed to the falsehood of *p*; for even though this act does not entail that one is committed to the falsehood of *p*, it is absurd to commit oneself to having such a commitment just when one also commits oneself to the truth of *p*" (Baldwin 2007, 88). For simplicity, and because this idea of a normative commitment to truth is not entirely clear (is it more than just a belief that p is true?), we will consider a parallel explanation which attempts to explain Moorean absurdity explicitly in terms of TRUTH NORM.

3.1 Does TRUTH NORM help?

TRUTH NORM, recall, states:

> **TRUTH NORM**: It is a conceptual truth about belief that it is correct to believe a proposition if and only if it is true.

Suppose that this is correct. Can it be used to explain what's absurd about accepting MP? Our argument will be that it doesn't – it is vulnerable to our DILEMMA.

Let's begin however by thinking about how TRUTH NORM *might be thought to* explain the absurdity of accepting MP. If we assume TRUTH NORM, then one who accepts MP will be committed by its first conjunct to: [It is correct to believe that p]. So, if we assume TRUTH NORM, then one who accepts MP is committed to:

> MP*: It is correct to believe that p, but I don't believe that p.

So, if TRUTH NORM is true, then the Moore-paradoxical sentence, MP, commits one who accepts it to claiming - simultaneously - that it would be correct to do (or be doing) something, and that one isn't doing that thing. Perhaps *that's* what's absurd about accepting MP: it's absurd to simultaneously accept that it is correct to do (or be doing) something, and that one isn't doing that thing.

Does this really help to explain what's strange about accepting MP? An obvious response is: no, it's not really strange – certainly not incoherent, absurd or impossible - to judge that one is under a prescription to respond in some way, but that one isn't responding in that way. Think about ordinary cases in which the prescription takes the form of an 'ought' fact rather than a correctness fact. For example, suppose I judge: [I ought to give money to charity, but I don't]. In making this judgment, I don't seem to be saying anything strange or absurd and certainly not something *impossible*. It seems like a rather normal thing to judge. In fact, I judge it frequently.[6] Contrast it with: [Climate change is happening, but I don't believe it is happening]. This *really does* seem strange. There is a big difference, in that respect, between these two judgments. So, one might think, it is mistaken to attempt to explain the absurdity of accepting Moore-paradoxical sentences by using TRUTH NORM to change MP into MP*.

Perhaps, however, this is too simple. Perhaps a defender of TRUTH NORM would do better by supplementing the basic structure of explanation (i.e. that MP commits one to MP*) in some way. How could we do that? One might try to claim that there is something *irrational*

---

[6] It will be helpful here to forestall an objection. We are not claiming here that there is nothing strange, absurd, or irrational about the *situation* in which in which I judge that I ought to give charity, but don't do so. We are claiming that there is nothing necessarily strange, absurd, or irrational about a *judgement* with the content [I ought to give money to charity, but I don't], because it might be a psychologically astute judgement about my moral failings. We show how this distinction matters, and undermines any argument for TRUTH NORM, in section 3.3. (We are grateful to Christos Kyriacou in pressing us to clarify this point).

about judging MP*, and hence MP. This strategy seems to have something to recommend it. It's easy to suspect that there is something irrational in accepting MP*. And if in fact it *is* irrational, then we could explain the absurdity of accepting MP*, and hence MP, in these terms. it would be *impossible for a rational person* to judge MP* and hence MP. TRUTH NORM would have allowed us to explain the absurdity of accepting MP* and hence MP.

There are, however, major problems with this strategy, two of which we will detail. The first problem is that this is the wrong kind of explanation of the absurdity of judging MP. The second, more serious problem, is that it doesn't work as an explanation of the absurdity of judging MP anyway because it fails on our DILEMMA.

3.2 First Problem: Wrong Kind of Explanation

The first problem is really a preliminary. Suppose that it *is rationally impossible* to accept MP. The problem is that it isn't obvious that this is a satisfactory explanation of the absurdity of accepting MP. It doesn't yield the *right kind* of absurdity.

Why not? Simply put: If the problem with believing MP is that it is rationally impossible, then we should expect someone irrational to be able to do it. But I'm not sure that we do expect this. It is not at all obvious that it is possible for an irrational person to believe [p, but I don't believe that p]. Believing [p, but I don't believe that p] is, we might think, more like *psychologically* impossible than rationally impossible. Even if you're irrational, you simply won't be able to hold this belief (though of course you may believe that you hold it).

It is useful here to refer back to the discussion of deliberative transparency in the earlier part of the paper. There we noted that it is unsatisfactory to explain deliberative transparency in such a way that it is *merely rationally impossible* to violate transparency in deliberation. It is not clear that this is the right kind of impossibility. It isn't obviously possible for an irrational person to violate deliberative transparency, nor would it obviously be irrational should someone for whom it is possible violate it. Much the same, I think, is true with Moore-paradoxical sentences.

3.3 Second Problem: Applying our Dilemma

The second problem is more serious. According to the view we are considering it is irrational to judge MP*. The problem, as noted above, is that it isn't obvious that it *is* irrational. Certainly, we wouldn't want to just *assert* that it is. We would need some reason or argument for thinking that it is. Ideally, we would want a general principle that explains why it is irrational. What might such a principle look like?

The obvious place to look in this area is to some form of *judgment internalism*. According to a simplified, classic form of judgment internalism (this was already referenced in the discussion of transparency above but we've specified the *scope* here for reasons that will be clear shortly):

(4) If you're rational, then: if you judge [it is correct to V], then you will be motivated to V.

This kind of claim is not uncontroversial, but it is often defended in some form or other. Let's suppose it is true. Let's suppose, furthermore, that there are no problems with V-ing standing for believing (as well as acting, as it does in standard formulations).

Could it be used to explain why it is irrational to judge MP*, and hence MP? We don't think it could. The problem is really one of *scope*. What we need to explain is why it is irrational to judge [It is correct to believe that p, but I don't believe that p]. That's a wide-scope judgment. But (4) won't help us to explain the irrationality of this wide-scope judgment. (4) would only help us to explain why it is irrational to judge [It is correct to believe p], and then not believe p. The irrationality attaches to not believing p given the narrow scope judgment [I ought to believe p].

Consider an example from the domain of action that can help to make this contrast intuitive. According to classical judgment internalists, it is irrational for one to judge [I ought to give money to charity], and then not give money to charity. It is a form of practical irrationality. But these same judgment internalists don't claim that it is irrational for one to make the wide-scope judgment [I ought to give money to charity, but I don't give money to charity]. Classical judgment internalism doesn't entail that this is irrational at all. It might just be a psychologically astute judgment. Similarly for MP. Classical judgment internalism doesn't seem to show that it would be irrational to make the wide-scope judgment [p, but I don't believe p]. Again, we can think about this in terms of the first horn of our DILEMMA. TRUTH NORM supplemented by (4) isn't a good explanation of the explanandum. It doesn't explain the absurdity of accepting Moore-paradoxical sentences at all.

This argument demonstrates that classical judgment internalism *doesn't* show that it is irrational to accept MP*, and hence MP. It also helps us to see what kind of principle we *would* need to do this. We would need a principle in which the irrationality attaches to the wide-scope judgment. The principle would have to look like this:

(5) If you're rational, then: you won't judge [It is correct to V, but I'm not V-ing].

If this general principle were true, then judging MP*, and hence judging MP, would be irrational. As a consequence, we would have some kind of explanation of the absurdity of accepting MP.

So far so good. The problem, however, is that, there's no very good independent reason to accept (5). As I noted above, it doesn't seem irrational to make this judgment in ordinary non-doxastic cases. One might judge: [I ought to give money to charity, but I don't given money to charity]. I'm not thereby be irrational, at least not obviously so. Certainly, there doesn't seem to be anything really strange about this judgment. It's nothing like judging that climate change is happening, but I don't believe it is happening. So this principle is under-motivated.

We can think about this in terms of the second horn of our dilemma. If we supplement TRUTH NORM with (5) we could, potentially, explain the absurdity of accepting MP. But (5) is an implausible claim about the relation between accepting prescriptions and acting on

them. So whether TRUTH NORM is supplemented with (4) or with (5), it fails to escape our dilemma. So TRUTH NORM doesn't explain the absurdity of accepting Moore-paradoxical sentences.

4. Conclusion

We have argued that attempts to use TRUTH NORM to explain deliberative transparency and Moore-paradoxical sentences fail on a common dilemma. As a consequence, we don't think that TRUTH NORM is well-supported. If we are right, then TRUTH NORM can't be the basis for a satisfactory middle-ground in meta-epistemology.

There is a more general point in the background. Both of the arguments for TRUTH NORM that we have considered – from deliberative transparency and Moore-paradoxical sentences - are really attempts to show that TRUTH NORM does important explanatory work in the philosophy of mind. More specifically, they are attempts to show that TRUTH NORM explains some mental phenomenon that relates what we can (or cannot) believe to the truth of the content of that belief. We think that this *kind* of argument is *bound* to fail. It is bound to fail because, if TRUTH NORM is to provide a uniquely good explanation of the mental phenomenon in question, it will have to make use of an overly strong claim about the relation between our accepting the prescriptions that supposedly constitute belief, and what we subsequently can or cannot believe. Because of this general structure, we are optimistic that our argumentative strategy will generalise. Specifically, we are optimistic that our argumentative strategy will undermine other attempts to show that constitutive norms on attitudes are justified by their role in explaining the impossibility of mental phenomena.

References

Archer, S. 2017. "Defending Exclusivity." *Philosophy and Phenomenological Research* 94 (1): 326–41.

Baldwin, T. 2007. "The Normative Character of Belief." In *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, edited by M. S. Green and J. N. Williams, 76–89. Oxford: Clarendon Press.

Buleandra, A. 2009. "Doxastic Transparency and Prescriptivity." *Dialectica* 63 (3): 325–32.

Bykvist, K., and A. Hattiangadi. 2007. "Does Thought Imply Ought?" *Analysis* 67 (296): 277–85.

———. 2013. "Belief, Truth, and Blindspots." In *The Aim of Belief*, edited by T. Chan, 100–122. Oxford: Oxford University Press.

Côté-Bouchard, C. 2016. "Can the Aim of Belief Ground Epistemic Normativity?" *Philosophical Studies* 173 (12): 3181–98.

Engel, P. 2013. "In Defence of Normativism about the Aim of Belief." In *The Aim of Belief*, edited by T. Chan, 32–63. Oxford: Oxford University Press.

Greenberg, A. 2017. "The Truth Norm Account of Justification." Thesis, University of Cambridge. https://doi.org/10.17863/CAM.18315.

Huemer, M. 2007. "Moore's Paradox and the Norm of Belief." In *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, edited by S. Nuccetelli and G. Seay, 142–57. Oxford: Clarendon Press.

McHugh, C. 2013. "Normativism and Doxastic Deliberation." *Analytic Philosophy* 54 (4): 447–65.

McHugh, C., and D. Whiting. 2014. "The Normativity of Belief." *Analysis* 74 (4).

Owens, D. 2002. "Epistemic Akrasia." *The Monist* 85 (3): 381–97.

Pettit, P., and M. Smith. 1996. "Freedom in Belief and Desire." *Journal of Philosophy* 93 (9): 429–449.

Railton, P. 1994. "Truth, Reason, and the Regulation of Belief." *Philosophical Issues* 5: 71–93.

Rosen, G. 2001. "Brandom on Modality, Normativity and Intentionality." *Philosophy and Phenomenological Research* 63 (3): 611–23.

Shah, N. 2003. "How Truth Governs Belief." *The Philosophical Review* 112 (4): 447–482.

Shah, N., and J. D. Velleman. 2005. "Doxastic Deliberation." *The Philosophical Review* 114 (4): 497–534.

Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.

Steglich-Petersen, A. 2006. "No Norm Needed: On the Aim of Belief." *The Philosophical Quarterly* 56 (225): 499–516.

———. 2013. "The No Guidance Argument." *Theoria* 79 (1): 279–283.

Svavarsdottir, S. 1999. "Moral Cognitivism and Motivation." *The Philosophical Review* 108 (2): 161–219.

Wedgwood, R. 2002. "The Aim of Belief." *Philosophical Perspectives* 16: 267–297.

Williams, B. A. O. 1970. "Deciding to Believe." In *Problems of the Self*, 136–51. 1973. Cambridge: Cambridge University Press.