

# TRANSPARENCY AND REASONS FOR BELIEF

Benjamin WALD

**ABSTRACT:** Belief has a special connection to truth, a connection not shared by mental states like imagination. One way of capturing this connection is by the claim that belief aims at truth. Normativists argue that we should understand this claim as a normative claim about belief – beliefs ought to be true. A second important connection between belief and truth is revealed by the transparency of belief, i.e. the fact that, when I deliberate about what to believe, I can settle this deliberation only by appeal to considerations I take to show  $p$  to be true. It is natural to think that there is a connection between these two features of belief, that the fact that believing for non-evidential considerations would be irrational can help to explain why it is impossible, and Shah and Velleman make exactly this argument. However, as I shall argue, we cannot explain transparency on the basis of a normative requirement on belief. For this explanation to work non-evidential considerations would have to fail to be reasons for belief, and we would have to be able to explain why we are unable to form beliefs on the basis of non-evidential considerations by appealing to the fact that they fail to be reasons for belief. However, while it is plausible that non-evidential considerations are not in fact reasons for belief, the explanatory picture is the other way around. Such considerations only fail to be reasons for belief because we are unable to form beliefs on their basis.

**KEYWORDS:** transparency, aim of belief, truth, reasons, non-evidential considerations

## Introduction

Belief has a special connection to truth, a connection not shared by other representational mental states such as imagination or supposition. We can see this connection in the fact that there is nothing out of the ordinary in saying “I am imagining that it is raining, but it isn’t raining,” but the Moore-paradoxical “I believe that it is raining, but it isn’t raining” is strikingly odd. One way that many philosophers have tried to cash the connection between belief and truth is in terms of the claim that, in some sense, belief aims at truth. However, there is substantial disagreement over the correct philosophical account of this claim. Beliefs, after all, are not themselves an agent who can have their own aims. According to normativists, the claim that beliefs aim at truth should be understood as a normative claim.<sup>1</sup> To say that belief aims at truth is to say that one

---

<sup>1</sup> Cf. Pascal Engel, “Doxastic Correctness,” *Proceedings of the Aristotelian Society* 87, (2013): 199-216, Pascal Engel, “In Defense of Normativism about the Aim of Belief,” in *The Aim of*

ought to hold a belief only if it is true, or that truth provides the standard of correctness for belief, for example.

Along with being the aim of belief, truth also has another interesting relation to belief. We can normally only form beliefs on the basis of considerations that we take to show the belief to be true. This is why, on proposing his wager as an argument for belief in God, Pascal goes on to recommend means by which one could bring oneself to actually form the belief that God exists. Accepting that belief in God is a good bet is not sufficient to bring about belief in God; you need to attend church, take communion, study the bible, and so on in the hopes of altering your evaluation of the truth of the belief in order to bring about this doxastic change. The situation is very different in cases where you become convinced that a consideration shows a claim to be true. In this case, we can form the belief directly, without adopting other means. This phenomenon has been called the transparency of belief. In some sense, the question of what to believe is transparent to the question of what is the case.<sup>2</sup> This explains why only evidential considerations, considerations that show the belief likely to be true in some way, can help us settle the question of what to believe.

It is natural to think that the aim of belief and the transparency of belief are related in some way. Both, after all, involve a special relationship between belief and truth. In particular, if we accept that the aim of belief should be understood in terms of a normative role for truth, then this looks like it should help us explain the transparency of belief. The fact that forming beliefs that aren't true is normatively forbidden may help us explain why it is impossible to do so, or at least to do so directly. Nishi Shah and David Velleman<sup>3</sup> argue for just such an explanatory relation between the aim of belief and the transparency of belief. In fact, this explanatory link provides the main argument for their version of normativism about belief. For normativism to explain transparency, the agent

---

*Belief*, ed. Timothy Chan (New York: Oxford University Press, 2013), 32-63, Ralph Wedgwood, "The Aim of Belief," *Philosophical Perspectives* 16, (2002): 267-297, Nick Zangwill, "Directions of Fit and Normative Functionalism," *Philosophical studies* 91, 2 (1998): 173-203, Nick Zangwill, "The Normativity of the Mental," *Philosophical Explorations* 8, (2005): 1-19.

<sup>2</sup> This notion of transparency is related to but distinct from that described by Gareth Evans in *Varieties of Reference* (Oxford University Press, 1982). Evans focuses on the relation between truth and belief when we are forming beliefs, rather than in coming to know what we believe. I discuss how we should understand Velleman's notion of transparency in more detail in section 2.

<sup>3</sup> See Nishi Shah, "How Truth Governs Belief," *The Philosophical Review* 112, 4, (2003): 447-482, Nishi Shah and J. David Velleman, "Doxastic Deliberation," *The Philosophical Review* 114, 4, (2005): 497-534.

must be aware of, or at least sensitive to, the normative requirements on belief. Otherwise, these normative requirements could not explain the psychological fact of transparency. This leads Shah and Velleman to argue that normativism is not (or not only) an independent normative truth, but part of the possession conditions for the concept of belief. In order to possess the concept of belief, Shah and Velleman argue, an agent must accept a normative claim, that beliefs are correct if and only if they are true. Thus, believers sophisticated enough to possess the concept of belief, and hence able to deliberate explicitly about what to believe, cannot help but be aware of the norm of truth, and this awareness can explain why beliefs formed through deliberation must be formed on the basis of considerations taken to be relevant to the truth of the belief.

However, tempting as it is, I do not think this explanatory strategy can ultimately be successful. We cannot explain transparency on the basis of a normative requirement on belief, not even if this norm is part of the possession conditions for the concept of belief. As I shall argue, for this explanation to work non-evidential considerations must fail to be reasons for belief. Furthermore, we must be able to explain why we are unable to form beliefs on the basis of non-evidential considerations by appealing to the fact that they fail to be reasons for belief. However, while it is plausible that non-evidential considerations are not in fact reasons for belief, the explanatory picture is the other way around. Such considerations only fail to be reasons for belief because we are unable to form beliefs on their basis. In other words, if we were able to form beliefs for non-evidential considerations, then such considerations would in fact count as perfectly valid reasons for belief. It is only our inability to actually believe for such reasons that prevents them from being reasons for us. And this shows that Shah and Velleman's strategy of explaining transparency in terms of a normative requirement fails.<sup>4</sup> Furthermore, it provides strong reason to doubt that any similar explanatory strategy could succeed. This removes the main support for Shah and Velleman's theory of the aim of belief. But it also potentially has wider consequences. It remains quite plausible that the aim of belief and the transparency of belief have some kind of explanatory relation, and indeed that transparency is explained by the aim of belief. But if normativism cannot explain

---

<sup>4</sup> For other criticisms of Shah and Velleman's position see Kathrin Glüer and Åsa Wikforss, "Against Belief Normativity," in *The Aim of Belief*, ed. Timothy Chan (New York: Oxford University Press, 2013), 80-99, Conor McHugh, "Normativism and Doxastic Deliberation," *Analytic Philosophy* 54, 4 (2013): 447-465, Andrei Buleandra, "Doxastic Transparency and Prescriptivity," *Dialectica* 63, 3 (2009): 325-332, Asbjørn Steglich-Petersen, "No Norm Needed: On the Aim of Belief," *The Philosophical Quarterly* 56, 225 (2006): 499-516, Asbjørn Steglich-Petersen, "Does Doxastic Transparency Support Evidentialism?" *Dialectica* 62, 4 (2008): 541-547.

Benjamin Wald

transparency, then this provides some reason to doubt that normativism is the correct account of the aim of belief.

## 1. Transparency

So what exactly is transparency? As Shah and Velleman put it,

The deliberative question whether to believe that  $p$  inevitably gives way to the factual question whether  $p$ , because the answer to the latter question will determine the answer to the former.<sup>5</sup>

In other words, when we deliberate about whether to believe  $p$ , we must settle our deliberation on the basis of exactly the same considerations that we would use to settle the question of whether  $p$ . Other considerations, although we might think about them and perhaps even wish we could form our belief on their basis, just do not settle the question of whether  $p$ , and so cannot settle the question of whether to believe that  $p$  either. We cannot, for instance, come to believe  $p$  because it would make us feel better, or because believing it would be good for our health, or because it would make our spouse happy.

This stands in stark contrast to how we deliberate about attitudes such as imagining or hypothesizing. We can decide to imagine that  $p$ , or hypothesize that  $p$ , for reasons that are utterly irrelevant to the truth of  $p$ . I can imagine that I have won an award just because imagining this would make me happy, but I cannot believe that I have won the award because the belief would make me happy. Thus, there is some special link here between belief and truth that shows up in our first personal deliberation about what to believe. It is important to note that this is not itself a normative claim – it is not that it is wrong to believe for pragmatic reasons, but that it is impossible to settle deliberation about what to believe by reference to anything other than evidential considerations, i.e. considerations we take to bear on the truth of the claim. There are actually two related claims being made here. The first is that one question, whether to believe that  $p$ , is transparent to a second question, whether  $p$ , when we deliberate. The second claim is that only what the agent takes to be evidence that  $p$  is true can be used by the agent to settle the question of whether  $p$ . However, I shall focus on the first claim, and take the second claim as given. The second claim will also gain some support from Shah and Velleman’s account of the nature of deliberation, discussed in section 3.<sup>6</sup>

---

<sup>5</sup> Shah and Velleman, “Doxastic Deliberation,” 499.

<sup>6</sup> The fact of transparency is not uncontroversial. Some philosophers think it is possible, and even sometimes rational, to give weight to non-evidential considerations in deliberating about what to believe (Cf. Conor McHugh, “The Illusion of Exclusivity,” *European Journal of*

Now, we should be clear about the strength of this transparency claim. Obviously, we might still be influenced in deciding whether to believe  $p$  by facts that have nothing to do with the truth of the belief. If I deliberate about whether to believe that my wife is cheating on me, I may be influenced by my deep desire not to believe this into discounting good evidence, in a way I would not have done had the question been about the faithfulness of someone else. What transparency rules out is that I could consciously decide not to believe that my wife is cheating on me on this basis. I can still be influenced by considerations that do not bear on the truth of the belief, but these must operate ‘behind the scenes,’ so to speak. The way in which these factors might influence my deliberation about whether or not to believe  $p$  is exactly the same way that they would operate in my deliberation about whether  $p$ , so we can retain the idea that the first question is transparent to the second.<sup>7</sup> Furthermore, the claim that deliberation operates in this way is not merely armchair philosophical speculation. Psychological research on cases of so-called ‘motivated reasoning,’ where reasoners are incentivized to come to particular conclusions, suggests that the influence of practical incentives is indirect. Studies shows that, while people are in fact more likely to form a belief when they have been given practical incentives to form that belief, there is no conscious link between the non-evidential considerations and the formation of the belief. People spent longer looking at evidence that supported the belief they were incentivized to form, and spent longer searching their memory for instances that supported the desired belief,<sup>8</sup> and the subjects were presumably unaware of this bias in their search for evidence. This provides empirical support for the claim that we can only form beliefs based on evidential considerations – when practical considerations affect our judgment, they do so by subconsciously affecting the way we look for or deliberate on evidential considerations, rather than by figuring explicitly in our deliberation. It is plausible that this is not just a contingent limitation on human believers, either. After all, a being that could form beliefs on the basis of non-evidential considerations could form beliefs on the basis of considerations they knew to be irrelevant to the truth of the belief. Thus they

---

*Philosophy*, forthcoming, DOI: 10.1111/ejop.12032). However, I shall assume for this paper that transparency is a real phenomenon.

<sup>7</sup> We can also retain the claim that only considerations taken to be evidence for  $p$  can be used to settle the question of whether  $p$ , since the non-evidential factors are not being taken by the agent to settle the question of whether  $p$ , but instead unconsciously influencing the agent’s thinking about whether other considerations are good evidence for  $p$ .

<sup>8</sup> Cf. Arie W. Kruglanski and Donna M. Webster, “Motivated Closing of the Mind: ‘Seizing’ and ‘Freezing,’” *Psychological Review* 103, 2 (1996): 263-283, Ziva Kunda, “The Case for Motivated Reasoning,” *Psychological Bulletin* 108, 3 (1990): 480-498.

Benjamin Wald

could form beliefs without regard to the truth of the belief, and such doxastic voluntarism is generally taken to be conceptually impossible,<sup>9</sup> making transparency a conceptual truth about belief.<sup>10</sup>

## 2. Belief as a Normative Concept

Shah and Velleman argue that their version of normativism provides the best explanation of the phenomenon of doxastic transparency. They argue that the concept of belief has as part of its possession conditions the acceptance of a normative claim: namely, that beliefs are correct only if they are true.<sup>11</sup> Thus, to possess the concept of belief at all requires us to endorse a normative claim about when it is correct to hold a belief, so no believer can fail to be aware of this normative claim. The fact that believers necessarily endorse a norm for belief can be used, Shah and Velleman argue, to explain doxastic transparency.

Transparency, as Shah and Velleman understand it, only shows up when we deliberate about what to believe. As I said above, it is possible for belief to be influenced by non-evidential considerations, as long as these considerations operate behind the scenes. When beliefs are formed without deliberation, however, all of the influences on belief are similarly behind the scenes. It is only in deliberation that we explicitly consider what considerations count as reasons for forming a belief, and hence only here that there is a difference between the role of some considerations as reasons on which the belief is formed as opposed to mere causal influences in the formation of belief. Thus there is no distinction between the way that evidential and non-evidential considerations operate on beliefs formed without deliberation. However, when we deliberate explicitly about what to believe, only evidential considerations are relevant to settling the question. So what explains the fact that transparency shows up only when we deliberate? Well, because the deliberation is about what to *believe*, the agent

---

<sup>9</sup> Jonathan Bennett, "Why is Belief Involuntary?" *Analysis* 50 (1990): 87-107, Bernard Williams, "Deciding to Believe," in *Language, Belief, and Metaphysics*, ed. Howard E. Kiefer and Milton K. Munitz (New York: SUNY Press, 1970), 95-111.

<sup>10</sup> If transparency has a contingent psychological explanation then so much the worse for attempts to give a normative explanation of the phenomenon. However, I shall assume, in line with Shah and Velleman, that transparency is a conceptual truth.

<sup>11</sup> 'Correct' here is supposed to be a normative term, rather than a purely descriptive term. Several philosophers take the norms of belief to be given in terms of correctness. See, for example, Wedgwood, "The Aim of Belief," Engel, "In Defense of Normativism," and Alan Gibbard, "Truth and Correct Belief," *Philosophical Issues* 15 (2005): 338-351. However, nothing in the argument hinges on using correctness: for our purposes the result is the same if the norm of belief is given in different normative terms, such as what we ought to believe, instead.

necessarily applies the concept of belief in thinking about the outcome of deliberation. If Shah and Velleman are right about the possession conditions for this concept, then this entails that the agent endorses a norm that says that the belief which is the outcome of the deliberation will be correct if and only if it is true. Furthermore, for some mental process to count as deliberation, the agent must aim to reach the correct conclusion. A mental activity that was not aimed at reaching the correct outcome wouldn't count as deliberation at all – it would be idle imagining, perhaps, or even just a disconnected series of thoughts. So, in deliberating about what to believe, we are aiming to form a correct belief, due to the nature of deliberation, and we accept that only true beliefs are correct, due to the nature of belief. This, Shah and Velleman argue, shows that we are committed to forming the belief based only on factors we take to be relevant to its truth – just what doxastic transparency requires.<sup>12</sup>

However, I do not think this proposed account could truly explain transparency. To see why, consider what the strength of the proposed norm would have to be for it to explain transparency. Normally, the fact that a norm applies to something does not serve to constrain deliberation in the way that transparency does. Imagine that a friend asks me what I think of their haircut, and I judge that the new look is a colossal mistake, so I deliberate about what to say. Imagine further that I endorse a norm that forbids lying. Nonetheless, it seems that my deliberation could still include considerations such as the fact that telling the truth will hurt their feelings, and I might well end up choosing to act on this consideration, despite my acceptance of the norm against lying. The norm forbidding lying tells me that I should not say that  $p$  unless I think  $p$  is true, but this does not prevent me from taking into account or acting on considerations that have nothing to do with the truth of  $p$ . With transparency, on the other hand, these other considerations are prevented from having any influence. Thus, if transparency is explained by a norm, this norm must be of a special sort, unlike familiar norms such as the one forbidding lying.

We might try claiming that the norm of belief, unlike the norm against lying, is a *decisive* norm. While the norm against lying may provide some reason against lying, this reason still needs to be weighed against competing reasons to see if it is the strongest reason in this instance. However, perhaps the reason provided by the norm of belief is guaranteed to always be a decisive reason, outweighing any competing reasons. Thus, the agent has no need to consider

---

<sup>12</sup> Shah and Velleman also discuss in more detail their conception of the nature of deliberation and a mechanism for how we transition from deliberation to judgment and from there to belief, but the details of this account are not relevant to my criticism of it, so I omit them here.

other potential reasons for belief, since they can be sure that the reason provided by the norm of belief will always win out. However, this too falls short of accounting for transparency. Even if an agent knows that some consideration provides a decisive reason, it still seems possible for them to be swayed in their deliberation into acting for a different reason. Acting against what one takes to be a decisive norm is irrational, to be sure, but it is also a familiar phenomenon – if this weren't possible, then weakness of will would be much less prevalent. Consider again the norm against lying. Perhaps I have read a lot of Kant recently, and come to endorse the view that the norm against lying is a decisive norm, never outweighed by competing considerations. It still seems perfectly possible that, in a particular case, I might end up, through weakness of will, considering the harm to my friend's feelings, and acting on this basis. But in the case of doxastic deliberation, such weakness of will is not just irrational, but impossible. Not even a decisive norm seems to explain this impossibility.<sup>13</sup>

A final, and more promising, option is to hold that the norm of belief is a *silencing* norm. On this view, the norm of belief not only provides reasons that outweigh any competing reasons, it silences competing reasons, prevents them from having any rational weight at all. This entails that non-evidential considerations will fail to be reasons for belief. After all, the norm of belief is always in operation, and hence it will always silence non-evidential considerations. If they are always silenced, then non-evidential considerations will never have any weight in any deliberation about what to believe. But a consideration that never has any weight is thereby not a reason at all, so non-evidential considerations will not count as reasons for belief. Alternatively, we might think that, rather than the norm of belief silencing and hence eliminating competing reasons, there just never were any other reasons in the domain of belief in the first place. Perhaps the norm of belief provides the only reasons to be had when it comes to belief. These two explanations are structurally distinct, but the upshot is the same in either case – non-evidential considerations just do not count as reasons for belief.

If this were the correct interpretation of the strength of the proposed norm of belief, then Shah and Velleman would be arguing that possessing the concept of belief requires us to hold that the only things that count as reasons for belief at all are evidential considerations. This has better prospects of explaining why it is impossible, not just irrational, to form a belief on non-evidential considerations

---

<sup>13</sup> This problem for Shah and Velleman's account has been previously noted by Steglich-Petersen, "No Norm Needed" and Sergio Tenenbaum, "Knowing the Good and Knowing What One is Doing," *Canadian Journal of Philosophy* 35 (2012): 91-117.

when deliberating. To believe on the basis of a non-evidential reason would, on this interpretation, involve deliberately forming a belief not just on the basis of a less pressing reason, but on the basis of something that is not even thought to be a reason at all. It seems plausible that this is not in fact possible. In the case of practical reason Joseph Raz<sup>14</sup> points out that I cannot choose to have a coffee because I love Sophocles. If my love of Sophocles fails to in any way render my drinking coffee intelligible, i.e. fails to be a reason to drink coffee, then this consideration cannot be my reason for acting, and so cannot settle my deliberation about what to do. Similarly, if a candidate reason for believing *p* wouldn't render the formation of that belief at all intelligible, then it is plausible that it cannot be the agent's reason for believing *p*. The belief might be caused by the consideration, through some arational psychological process, but unless the consideration is seen as at least some reason for the belief, it couldn't count as the agent's conclusion in deliberation.

Furthermore, it seems plausible that this in fact how Shah thinks of the norm as functioning. For instance, he says that the effect of endorsing the norm of truth is that a strong disposition to block the influence of non-evidential types of influence is activated in cases of belief-formation that are governed by an agent's application of the concept of belief.<sup>15</sup> This suppression of non-evidential considerations sounds more like a case of silencing such considerations than it does merely outweighing them. Similarly, Shah states that

belief's standard of correctness does determine what counts as a reason for belief from within the first-personal deliberative point of view.<sup>16</sup>

Shah here claims not only that the standard of correctness *provides* a reason, but also that it determines *what counts as a reason*. In other words, the claim of any other consideration to count as a reason at all depends on the norm of correctness, which suggests that it silences competing reasons.

This approach obviously requires that there in fact be no non-evidential reasons for belief. Furthermore, the argument requires that our inability to form beliefs for non-evidential reasons were explained by the non-existence of any such reasons. I shall argue in section four that, were we able to believe for non-evidential reasons, some of them would be perfectly good reasons for belief. This may seem to commit me to the unpopular view that there are in fact non-

---

<sup>14</sup> Joseph Raz, "When We Are Ourselves: The Active and the Passive," in his *Engaging Reason: On the Theory of Value and Action* (New York: Oxford University Press, 2002), 5-21.

<sup>15</sup> Shah, "How the Truth Governs Belief," 473.

<sup>16</sup> Shah, "How the Truth Governs Belief," 472.

evidential reasons for belief. While this view is not obviously false,<sup>17</sup> I am not committed to accepting this conclusion. Instead, we can hold that there are no non-evidential reasons for belief, but that the explanation of this fact is that it is impossible to form beliefs on the basis of non-evidential considerations. This, combined with a modest internalism about reasons for belief, entails that non-evidential considerations cannot be reasons for belief. However, this order of explanation will not help Shah and Velleman's argument, since they need the opposite order of explanation. In section five, then, I shall present a brief explanation of weak internalism about reasons for belief and show how, combined with transparency, it entails that there are no non-evidential reasons for belief, and also why this nonetheless is no help for Shah and Velleman. This will show how we can accept the arguments of section four without being committed to the existence of non-evidential reasons for belief.

### 3. Non-Evidential Reasons for Belief

For something to be a reason for belief implies that were we to form a belief on the basis of this consideration we would not be rationally criticisable, *ceteris paribus*,<sup>18</sup> whereas we are rationally criticisable for forming beliefs on the basis of things that are not in fact normative reasons for belief. Thus, if we accept Shah and Velleman's claim that all believers must accept a norm that silences all non-evidential reasons for belief, then it should be rationally criticisable to form beliefs on the basis of these considerations. In this section, I shall argue that this is not true – if we were to form beliefs on the basis of some non-evidential considerations, this would not be rationally criticisable. Now, as I shall argue in section five, we might still hold that these considerations fail to be reasons for belief. In particular, they may fail to be reasons for belief precisely because we cannot form beliefs on the basis of such considerations. However, even so, it still remains true that were we able to form beliefs for these reasons, there would be nothing rationally criticisable about doing so.

---

<sup>17</sup> For a defense of non-evidential reasons for belief, see Andrew Reisner, "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem," *Philosophical Studies* 145, 2 (2009): 257-272.

<sup>18</sup> The *ceteris paribus* clause here is important. We can be rationally criticisable for forming a belief on the basis of a genuine reason for belief if, for instance, there are stronger reasons against the belief, or a defeater for this reason is present. Still, there remains an important conceptual link between reasons for belief and rationality, which we can use to determine when a consideration counts as a reason for belief.

So why think that there are cases where it is not rationally criticisable to form a belief for non-evidential considerations? The basic argument is as follows. Opponents of non-evidential reasons for belief in general accept that non-evidential reasons can give us reasons to *bring it about* that we believe the proposition in question. For example, if an evil demon threatens to destroy the world unless you believe that the earth is flat, this provides you with a strong reason to take whatever means you can to bring it about that you believe this - read flat-earth arguments, try to convince yourself that there is a conspiracy against flat-earthers, get someone to hypnotize you, and so on. However, it seems very strange to say that it is rationally permissible to bring it about that you believe something, but were you able to bring yourself to believe the proposition directly, you would be rationally criticisable for doing so. Imagine someone who has the capacity to form beliefs on the basis of both evidential and non-evidential considerations, and who is deliberating about what to believe in the evil demon scenario. It seems highly implausible that he would be rationally criticisable for forming the belief that the world is flat on the basis of the demon's threat. Of course, we might object that forming the belief directly is not criticisable but impossible. I think this is exactly right, and perhaps, as I suggest in section five, we might think that as a matter of fact we would therefore have no reason to do as the demon commands, since it may be necessary for something to be a reason for belief that it is possible to form beliefs for this very reason. However, Shah and Velleman cannot appeal to this impossibility without rendering their position circular. Shah and Velleman need it to be impossible to form beliefs for non-evidential reasons because the agent would see it as violating the norm on belief that they must endorse to count as a believer; they cannot then explain the fact that it would violate the norm in terms of it being impossible to form the relevant belief. This is the core of my argument that it would not always be rationally criticisable to form a belief on the basis of a non-evidential consideration, and thus that there cannot be a general silencing norm forbidding forming beliefs for such reasons.

The counterfactual with which I frame my argument here may seem problematic. I argue that in some circumstance if we were able to form beliefs for non-evidential considerations then this would not be rationally criticisable. However, transparency is a conceptual truth, and hence necessary, and it states that we can only form beliefs for evidential considerations. So, the antecedent of this counterfactual is necessarily false, and thus according to the standard Lewisian semantics for counterfactuals the whole counterfactual is vacuously true. Similarly, the existence of an agent who could believe for non-evidential reasons

is similarly impossible, making the counterfactual framed in terms of such an agent also vacuous and uninformative.

I am not convinced that all counterfactuals with impossible antecedents are in fact vacuous. Consider, for instance, the claim that if Pythagoras' theorem were false, mathematicians wouldn't believe it. This seems non-vacuously true despite the necessary falsity of the antecedent.<sup>19</sup> Still, I think we can make the same point without appealing to such counterpossible scenarios. Imagine instead an agent who is unsure of whether or not they can believe on the basis of non-evidential considerations. Perhaps they think they probably can't, but they aren't sure. When confronted by the evil demon considered above, this agent therefore tries to believe on the basis of the non-evidential considerations. If it is rationally criticisable to believe for non-evidential reasons, then it is rationally criticisable to try to believe for these reasons. But this agent does not seem rationally criticisable for making this attempt (although, depending on the scenario, they may be rationally criticisable for failing to recognize that it is impossible). On the other hand, imagine that this agent instead tries to believe that the earth is flat in order to annoy his philosophy teacher. In this case, the attempt does seem rationally criticisable, in a way that the attempt to satisfy the demon is not. While the goal is impossible in both cases, in the former the consideration speaks in favour of the belief, while in the latter it fails to do so. However, on Shah and Velleman's account, all such non-evidential considerations should equally be rationally criticisable to base beliefs on (or to try to do so), since all are equally ruled out as reasons by the aim of belief.

We might worry about the principle that, if it would be rationally criticisable to do something, then it is rationally criticisable to try to do it. But this principle is suggested by the plausible claim that agents are not rationally criticisable for failures caused by external factors over which they have no control. If I intend to visit Paris, I am rationally criticisable if I fail to buy a ticket or don't make plans to arrive at the airport on time, but I am not criticisable if the flight is cancelled due to a surprise storm. Once we subtract factors over which the agent has no control, however, trying one's hardest to do something and actually doing it are identical. To try one's hardest to do something is to do everything in one's

---

<sup>19</sup> For further discussion of the view that counterpossibles can be non-vacuously true or false, and how to provide a semantics for them, see Jens Christian Bjerring, "On Counterpossibles," *Philosophical Studies* 168, 2 (2014): 327-353, Berit Brogaard and Joe Salerno, "Remarks on Counterpossibles," *Synthese* 190 (2013): 639-660, David Vander Laan, *Lewisian Themes: The Philosophy of David K. Lewis* (New York: Oxford University Press, 2004), Daniel Nolan, "Impossible Worlds: A Modest Approach," *Notre Dame Journal of Formal Logic* 38, 4 (1997): 535-572.

power to bring it about, and if external circumstances cooperate then one succeeds - there is no rationally relevant gap between the attempt and the success. This strongly suggests that the rational status of a successful action should be the same as a sufficiently determined attempt.

A defender of Shah and Velleman might object that I have not established the direction of explanation I am arguing for between reasons and belief. I have argued that it is our inability to believe on the basis of non-evidential reasons that explains their not being reasons, rather than the other way around. But, it could be objected, the counterfactuals I have made use of don't necessarily show this, even if true. They might instead show that if we were able to believe on the basis of these considerations that would be because they would, in that counterfactual situation, be reasons. Thus, a defender of Shah and Velleman's view could object that if I could respond to non-evidential considerations then they would be reasons, but not because my inability to respond explains their not being reasons. Instead, if I could respond to them, that would be because they were reasons.<sup>20</sup>

However, it does not seem open to Shah and Velleman to claim that the considerations at issue are reasons in the counterfactual scenario described. Shah and Velleman are committed to the claim that it is a conceptual truth that only evidential considerations can be reasons for belief. Thus, in the scenario presented earlier, where an agent believes that the earth is flat in order to prevent the demon destroying the world, they would need to claim that the agent takes the demon's threat to be evidence that the world is flat. But it seems clear that the demon's threat is not evidence, and we can even add to the scenario that the agent doesn't take it to be evidence, and still generate the intuitive judgment that the agent's believing that the earth is flat on this basis would not be rationally criticisable. Thus, Shah and Velleman cannot offer as an explanation for the counterfactual the claim that the demon's threat is a reason in this scenario, since this would be to abandon the claim that reasons must, as a matter of conceptual necessity, be considerations the agent takes to be evidence.

Cases of 'motivated irrationality' might seem to provide examples of cases where it is in fact rationally permissible to do indirectly what it would be rationally criticisable to do directly. Parfit,<sup>21</sup> for example, imagines a scenario in which a robber is trying to force you to open your safe so he can steal the gold, and he is willing to torture you or threaten your family in order to get you to comply. If you had a pill that would make you utterly irrational, Parfit argues, then the rational thing to do would be to take the pill. After all, if you were

---

<sup>20</sup> Thanks to an anonymous referee for pressing me on this point.

<sup>21</sup> Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984).

utterly irrational, you would not respond rationally to the robber's threats, and so he would realize that there was no point in making these threats or carrying them out. Parfit describes the scenario playing out:

Reeling about the room, I say to the man: 'go ahead. I love my children. So please kill them.' The man tries to get the gold by torturing me. I cry out: 'this is agony. So please go on.' Given the state that I am in, the man is now powerless. He can do nothing that would force me to open the safe. Threats and torture cannot force concessions from someone is so irrational. The man can only flee, hoping to escape the police.<sup>22</sup>

Cases of threats and deterrence may provide similar examples. If I can inculcate in myself a disposition to always carry out my threats, even when doing so is irrational,<sup>23</sup> this may be beneficial. Those I threaten, aware of my irrational disposition, might then accede to my demands, and thus I never have to actually carry through on my threats, so I end up benefitting.<sup>24</sup> In these situations it seems perfectly rational to make oneself irrational. However, performing the irrational actions directly would still be rationally criticisable. I would be rationally criticisable to directly act on my terrible threat, even if it is rational to bring about my disposition to carry out threats. This seems to provide a counter-example to the above argument, by suggesting cases in which it is rationally permissible to bring about what would be irrational to do directly.

However, these cases are importantly different from the scenario we began with. In the cases of motivated irrationality, what is rational to bring about is the *disposition* to perform irrational acts. We are not seeking to indirectly bring about an attitude or an action, but a disposition, and the benefit of the indirect actions is derived from the benefit of having this disposition. Performing the irrational actions directly would fail to realize this benefit. Carrying out a threat out of the blue fails to achieve the benefit of having the disposition to carry out threats, since the whole point of the disposition is deterrence. In the evil demon case we began with, the situation is different. What renders it rational to bring it about that I believe the earth is flat is the benefit of believing that the earth is flat; in particular, the fact that this belief will persuade the demon not to destroy the world. But forming the belief directly also achieves this very same result. It would be very odd if one and the same result could be achieved either directly or

---

<sup>22</sup> Parfit, *Reasons and Persons*, 13.

<sup>23</sup> Due to the amount of harm carrying out my threat will lead to both for myself and for the threatened individual.

<sup>24</sup> Cf. Parfit, *Reasons and Persons*, David Gauthier, "Assure and Threaten," *Ethics* 104, 4 (1994): 690-721.

indirectly, and the benefit of the result makes the indirect method rational but not the direct method. It is not as if the direct method has terrible side effects that the indirect method lacks - the end result is identical, and if anything the indirect methods are more liable to produce undesirable side effects and cost extra time and effort.

But what if the evil demon threatens to destroy the world unless you believe something irrational? Wouldn't this show that sometimes one has a reason to indirectly bring something about, even though doing it directly would be irrational? After all, I have very good reason to bring it about that I believe something irrational, but by hypothesis I don't have any good reason to believe the irrational thing directly; if I did, it would be rational, and hence useless in my attempt to satisfy the demon. However, this is actually just another case where it is independently impossible to form a belief for a given reason; and hence, it has the wrong order of explanation to help Shah and Velleman. This is a bit easier to see in the practical case. Imagine that the demon has instead threatened to destroy the world unless I perform an irrational action. Imagine that, in order to comply with the demon, I hit myself in the head with a hammer as hard as I can. Is this irrational? It certainly would be normally. But in this case, if I am doing it because I believe this is the best way to prevent the demon destroying the world, then it in fact seems perfectly rational. Which, of course, defeats the point. So we have a conundrum. Almost anything I could do would be rendered rational by seeing it as a means to preventing the world being destroyed. And anything I shouldn't do even to save the world presumably still shouldn't be done. If I have a button that destroys the galaxy, then pressing this in order to prevent the world being destroyed would still be irrational, but only because pressing it is so much worse than the world being destroyed, so someone who does so is rationally (and morally) criticisable.<sup>25</sup> Thus, it seems that the only way to do something irrational is for that action not to be done in order to satisfy the demon, since this is a strong enough reason to render almost anything rational. But it seems plausible that the only way to do something irrational without it being done in order to save the world is to bring about the irrational action indirectly, perhaps by inculcating an irrational disposition in oneself and trying to forget the demon's threat altogether. However, this is not because there is anything wrong with the reasons I would be acting on if I acted directly. Preventing the demon from destroying the world is an

---

<sup>25</sup> The same goes for non-consequentialist reasons, although it is hard to think of examples of non-consequentialist reasons that are stronger than the reason in favour of saving the world. Still, if they exist, then presumably it would be wrong to violate this requirement even to save the world, so someone who does so is still rationally criticisable.

excellent reason to do something irrational, or would be if I could actually act on it. Sadly, precisely because it is such a strong reason, it is impossible to act for this reason. It will render the action I am trying to perform rational after all, defeating the purpose. Thus, here too, the order of explanation is wrong for this example to help Shah and Velleman. It is not that I lack sufficient reason to act irrationally, and this explains my inability to so act. Instead, only my inability to act for this reason prevents it from being an excellent reason. The same goes for the case of belief.

One could object at this point that there is at least some sense in which someone who intentionally forms a false belief in order to save the world is rationally criticisable. Even if we recognize the great practical benefits at stake, we might still say that they would be *epistemically* irrational to form a belief they took to be false.<sup>26</sup> The idea here is that epistemic reasons and practical reasons are not commensurable; they are two entirely separate standards of assessment. Epistemic reasons, on this view, are just those reasons that have to do with the truth or falsity of our beliefs, and we are epistemically irrational insofar as we fail to believe in accordance with these reasons. We can, of course, label a certain class of reasons as ‘epistemic reasons,’ and define corresponding notions of ‘epistemically rational’ and so on to accompany it. But this fails to address the main issue. The question of what to believe is a deliberative question facing agents. Recall our imagined agent who can deliberatively form beliefs on the basis of either evidential or non-evidential considerations. Such an agent would need to determine what they should believe. To tell them that there is one answer to what they should epistemically believe, and a different answer to what they should practically believe would be unhelpful- they would still be left with the unanswered question of what they should believe *simpliciter*. Imagine such an agent who is confronted by the evil demon who will destroy the world unless she believes the earth is flat. She knows the practical reasons favour believing that the world is flat, and the epistemic reasons favour believing that it is not flat, but she remains unsure what to believe. Is there really no further fact of the matter about what she should believe? This seems highly implausible.<sup>27</sup> Of course, we might

---

<sup>26</sup> It's not obvious that this solution will help Shah and Velleman. However, perhaps they could provide some further argument to explain why in deliberation we need to form beliefs not just for a reason, but for an epistemic reason. As I shall show, this move does not seem promising.

<sup>27</sup> This same point is frequently made about practical rationality. Some theorists about practical reason claim that there is no such thing as what ought to be done *simpliciter*, but only what ought to be done according to morality, what ought to be done according to self-interest, and so on (Cf. Philippa Foot, “Morality as a System of Hypothetical Imperatives,” *The Philosophical Review* 81, 3 (1972): 305-316, David Copp, “The Ring of Gyges: Overridingness and the Unity of

object that a believer who is capable of settling deliberation for either evidential or practical reasons is incoherent. The mental states formed by such a being just would not count as beliefs, we might suspect. But if so, this just shows that the true explanation of transparency lies with the explanation of why positing such a believer is incoherent, rather than with normativism about the aim of belief.

#### 4. Internalism about Epistemic Reasons

The above argument suggests that there is nothing rationally criticisable about believing for non-evidential reasons. However, it does not necessarily follow that such considerations are in fact reasons for belief. Consider the following principle:

EPISTEMIC REASONS INTERNALISM: For some consideration  $p$  to be a reason to believe  $q$ , it must be possible for an agent to believe  $q$  for this very reason.

To unpack this claim, let us introduce the idea of a motivating reason for belief, by analogy with concept of a motivating reason for action. The motivating reason for one of my beliefs is, roughly, the consideration in light of which I form the belief, and also what I would appeal to if my belief were challenged.<sup>28</sup> Note that the way that I have described it, a motivating reason for belief is not usually a psychological fact, but instead a fact, or putative fact, about the world. After all, I would not usually appeal to my own psychology if challenged to defend one of my beliefs – I would appeal to what I took to be evidence for the belief.<sup>29</sup> We can

---

Reason,” *Social Philosophy and Policy* 14, 1 (2007): 86-106). However, many other theorists object that there must be such a thing as what we ought to do simpliciter, and that this is shown by the fact that, even after being told what morality recommends and what self-interest recommends, it is coherent and indeed natural to persist in asking what one ought to do (Cf. Stephen Darwall, “Reasons, Motives, and the Demands of Morality: An Introduction,” in *Moral Discourse and Practice: Some Philosophical Approaches*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (New York: Oxford University Press, 1997), 305-312, Sharon Street, “Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About,” *Philosophical Issues* 18 (2008): 207-228).

<sup>28</sup> This is not meant to be a definition of an explanatory reason for belief: it is intended to fix our attention on the appropriate phenomenon. For further discussion of motivating reasons in practical reason, see Michael Smith, *The Moral Problem* (Oxford: Blackwell Publishing Ltd. 1994), Jonathan Dancy, *Practical Reality* (Oxford: Clarendon Press, 2000), Kieran Setiya, *Reasons without Rationalism* (Princeton: Princeton University Press, 2010). For discussion of this same distinction applied to belief, see Kieran Setiya, “Epistemic Agency: Some Doubts,” *Philosophical Issues* 23 (2013): 179-198, Pamela Hieronymi, “The Wrong Kind of Reason,” *The Journal of Philosophy* 102, 9 (2005): 437- 457.

<sup>29</sup> Sometimes psychological facts may be motivating reasons for belief, as when I take the fact that I keep having sad thoughts as evidence that I am depressed, but this will not be the usual case.

contrast this with the idea of a normative reason for belief, which is the kind of consideration that actually counts in favour, normatively, of the belief. With this contrast in place, the suggestion under consideration is that for some fact to be a normative reason for belief, it must be capable of being a motivating reason for belief. This principle would be the theoretical analogue of a fairly weak form of internalism about practical reasons. Some philosophers hold that for something to be a normative reason for an agent that agent must be able to become motivated to act on the reason given their existing desires and psychology.<sup>30</sup> Our proposed principle is much weaker, since as I will argue the notion of possibility at issue is weaker than the kind of psychological possibility appealed to in these more strongly internalist arguments.

Epistemic reasons internalism may at first seem implausible. Consider someone who, perhaps due to very effective brainwashing in their youth, is unable to believe in the theory of evolution, and therefore unable to believe in the theory for the reason that it is supported by the best scientific evidence. This doesn't seem to show that this evidence thereby provides such a person with no reason to believe in the theory of evolution. The wealth of evidence for the theory still gives them very strong reason to believe it, even if they are unable to respond rationally to this evidence. However, I suspect that this is an issue of finding the correct notion of possibility. It may be psychologically impossible for the brainwashed individual to believe in the theory of evolution but this just shows that we should make use of a weaker form of possibility. The most plausible candidate is conceptual possibility. It must be at least conceptually possible for an agent to take some consideration as a motivating reason for belief for that consideration to be a normative reason for belief for that agent.

Furthermore, if we were right in claiming that transparency is a conceptual claim, then this will establish that non-evidential considerations cannot be reasons for belief. Of course, we will need some explanation of why transparency is a conceptual truth that does not, like Shah and Velleman's argument, rely on normativism. We might, for instance, argue that it is a conceptual truth about beliefs that they are mental states formed in response to evidential considerations. We can form representational mental states on the basis of considerations we do not take to bear on the truth of the content of such states, but these states will thereby fail to count as beliefs. They might instead be suppositions or

---

<sup>30</sup> Cf. Bernard Williams, "Internal and External Reasons," in his *Moral Luck*, (Cambridge: Cambridge University Press, 1981), 101-113, David Velleman, "The Possibility of Practical Reason," *Ethics* 106 (1996): 694-726.

imaginings. Pamela Hieronymi,<sup>31</sup> for example, presents a non-normative argument that, if successful, would also establish that it is a conceptual truth that beliefs must be held for evidential reasons, by identifying the belief that *P* with the agent's answer to the question of whether *p*. Non-evidential considerations may make me wish that I could answer this question one way rather than another, but I can only actually settle the question, and thus form the belief, on the basis of considerations I take to be relevant to whether *p*, i.e. evidence. Shah and Velleman's argument relies on the idea that "deliberation is reasoning aimed at issuing in some result in accordance with norms for results of that kind"<sup>32</sup>. They then go on to argue that it is the norm governed aspect of deliberation that explains transparency - any agent who possesses the concept of belief must accept that truth is the only norm for beliefs. But we can use the characterization of deliberation given by Shah and Velleman to explain transparency even if we reject normativism. Deliberation not only aims to accord with norms for the result produced, it also aims at actually producing the result. Thus, deliberation about what to believe aims to produce belief. But if belief is, as a conceptual matter, something that must be formed on the basis of evidence, then deliberation about what to believe will be restricted to evidential considerations, because taking account of any other kind of consideration could not actually produce belief. If true, this account would show that there are no non-evidential reasons for belief, but this would be explained by transparency (together with the conceptual truths about deliberation and belief), and so could not be used to explain transparency without circularity. Thus, the truth of evidentialism on its own is not enough to save Shah and Velleman's argument. They need it to be the case that we cannot form beliefs for non-evidential reasons *because* there are no such reasons. However, epistemic reasons internalism need not establish this direction of explanation. Even if true, it might instead establish that there are no non-evidential reasons for belief because we cannot believe based on them.

## Conclusion

Shah and Velleman's view, then, fails to adequately explain transparency. Non-evidential considerations would be reasons for belief, if we were able to form beliefs on the basis of such considerations. Thus, we cannot explain our inability to form such beliefs as a result of their not being reasons for belief - to do so would be circular. And this same argument suggests that the prospects for any

---

<sup>31</sup> Hieronymi, "The Wrong Kinds of Reasons."

<sup>32</sup> Nishi Shah, "A New Argument for Evidentialism," *The Philosophical Quarterly* 56, 225 (2006): 481-498.

Benjamin Wald

normative theory of the aim of belief being used as an explanation for transparency are dubious. This is, of course, a particular problem for Shah and Velleman's view. The purported ability of their theory to explain transparency served as the major argument for the view, so if this explanation fails the view is left largely unmotivated. However, I think this argument has implications for other views about the aim of belief. The original thought, that transparency has something to do with the aim of belief, remains highly compelling. Why, in deliberation, must our answer to the question of what to believe be resolved by our answer to the question of what is the case? Well, it seems plausible that it is because we are trying to form a belief in deliberating, and belief aims at truth. Absent an account of the aim of belief, this explanation is merely a sketch, but it seems to be on the right track. If, as I have suggested, we cannot explain transparency in terms of a norm of belief, then we will have to reject the suggestion that normativists can explain transparency by appeal to the aim of belief. Now, perhaps there is an explanation of transparency that has nothing to do with the aim of belief. I have certainly not said anything to rule out this possibility. But a theory that could account both for transparency and for the aim of belief seems like it would have a distinct advantage, and the inability of normativist understandings of the aim of belief to provide such a unified account is a mark against it.<sup>33</sup>

---

<sup>33</sup> I would like to give special thanks to Sergio Tenenbaum, whose extensive comments on multiple drafts of this paper were invaluable. I would also like to thank Phil Clark, Elena Derksen, Mark Fortney, Jennifer Nagel, Luke Roelofs, Andrew Sepielli, audiences at the CPA annual congress and the University of Toronto grad forum, and two anonymous referees for very helpful comments on the paper.