

# Transparency and Reasons for Belief

## **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 the 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; 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. Shah and Velleman [2005] make exactly this argument, and indeed they use the ability of their version of normativism to explain transparency as one of the main arguments in support of the theory. However, as I shall argue, this approach is mistaken. 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. Furthermore, 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. 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. Furthermore, it provides strong reason to doubt that any similar explanatory strategy could succeed. Furthermore, since 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, this provides some reason to doubt that normativism is the correct account of the aim of belief.

## **1 Introduction**

Belief clearly 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 the 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 the truth should be understood as a normative claim<sup>1</sup>. To say that belief aims at truth is to say that one 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 belief in God 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 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 cannot help but be aware of the norm of truth, and this awareness can explain why beliefs formed

---

<sup>1</sup>Cf. Wedgwood [2002], Zangwill [1998, 2005], Engel [2013a,b]

<sup>2</sup>This notion of transparency is related to but distinct from that described by Gareth Evans [1982] when he introduced the term- it 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 this form of transparency in more detail in section 2.

<sup>3</sup>See Shah [2003], Shah and Velleman [2005]

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

## **2 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" [Shah and Velleman [2005, 499]. 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.

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>4</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 beliefs 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 (Cf. Kunda [1990], Kruglanski and Webster [1996]), 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 who 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 could form beliefs without regard to the truth of the belief, and such doxastic voluntarism is generally taken to be conceptually impossible<sup>5</sup>. This suggests that transparency is a

---

<sup>4</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>5</sup>Cf. Bennett [1990], Williams [1970]

conceptual truth about belief<sup>6</sup>.

### 3 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>7</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 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>8</sup>

---

<sup>6</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>7</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 [2002], Engel [2013b], Gibbard [2005]). 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.

<sup>8</sup>Shah and Velleman [2005] 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.

However, I do not think this proposed account can 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 “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 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 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 friends 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>9</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,

---

<sup>9</sup>This problem for Shah and Velleman's account has been previously noted by Steglich-Petersen [2006] and Tenenbaum [2012]

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 is the correct interpretation, 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 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 [2002] 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, in the case of theoretical reason, if my forming the resulting belief wouldn't be made intelligible at all by the purported reason for belief, then it plausibly couldn't even count as the agent's forming a belief for this reason at all. 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 [2003] 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” Shah [2003, 473]. This suppression of non-evidential considerations sounds more like a case of silencing such considerations than it does merely outweighing them. Similarly, For instance, 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” Shah [2003, 472]. Shah here claims not only that the standard of correctness *provides* a reason, but 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. However, as I shall argue in the next section, it also requires a particular explanation of why there are no non-evidential reasons for belief. As I shall show, although it is plausible that there are no non-evidential reasons for belief, the reason why this is the case also shows that the fact that there are no non-evidential reasons cannot be used to defend transparency.

## **4 Internalism about Reasons for Belief**

One possible argument Shah and Velleman might adopt to establish the claim that

there are in fact no non-evidential reasons for belief is to argue that for something to be a reason for me to believe P, it must be possible for me to believe P for this very reason<sup>10</sup>. In fact, Nishi Shah makes exactly this argument in Shah [2006] as part of his argument for evidentialism<sup>11</sup>. To clarify this idea, let us introduce the idea of an explanatory reason for belief, by analogy with concept of an explanatory reason for action. The explanatory 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>12</sup>. Note that the way that I have described it, an explanatory 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>13</sup>. We can contrast this with the idea of a normative reason for belief, which is the kind of consideration that actually counts in favor, 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 an explanatory reason for belief. This principle would be the theoretical analogue of a fairly weak form of internalism about practical reasons,<sup>14</sup> and so I shall call this principle internalism about epistemic reasons.

This principle may at first seem highly 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. To establish this claim, we might 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 imaginations.<sup>15</sup>

<sup>10</sup>This is what I take to be the basis of Hieronymi's argument for rejecting the existence of non-evidential reasons for belief in Hieronymi [2005]

<sup>11</sup>Shah fails to recognize the fact that, taken together, his argument for evidentialism and his argument about the nature of belief contradict one another. The two arguments require opposite orders of explanation between our inability to believe for non-evidential reasons and the status of such considerations as reasons, as I argue in more detail below.

<sup>12</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 explanatory reasons in practical reason, see Smith [1994]; Dancy [2000]; Setiya [2010]. For discussion of this same distinction applied to belief, see Setiya [2013], Hieronymi [2005].

<sup>13</sup>Sometimes psychological facts may be explanatory 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.

<sup>14</sup>Bernard Williams [1981] and David Velleman [1996], among others, 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. 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 that Williams and Velleman appeal to.

<sup>15</sup>Shah and Velleman [2005] in fact make a conceptual claim quite like this in supporting the claim that

The claim that normative reasons for belief must be capable of being explanatory reasons for belief is a controversial one, and I have not tried to argue for it in detail. My interest in this claim is that it might seem to offer a promising argument in support of Shah and Velleman's explanation of transparency. Their view requires that non-evidential considerations not count as reasons for belief, and internalism about epistemic reasons would, if true, establish this claim. If internalism about epistemic reasons is false, then so much the better for my criticism. However, even if it is true, this does not get Shah and Velleman off the hook. They need it to be the case that we cannot form beliefs for non-evidential reasons *because* there are no such reasons. However, the internalism about epistemic reasons would not establish this direction of explanation. Even if true, this argument would instead establish that there are no non-evidential reasons for belief because we cannot believe based on them. On Shah and Velleman's account, doxastic transparency is supposed to be explained by the normativity of the concept of belief, so the account is trying to explain why we cannot form beliefs for non-evidential reasons. It would, therefore, be circular to invoke this very phenomenon in defense of their account. Thus, if I can show that there are non-evidential considerations that would be reasons for belief were we able to form beliefs on their basis, then this will show that Shah and Velleman's explanation fails.

## 5 Non-Evidential Reasons for Belief

So why think that non-evidential considerations would be reasons for belief, were we able to believe based on them? The basic argument is as follows. 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 criticizable, *ceteris paribus*<sup>16</sup>. Now, the 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 criticizable for doing so. This suggests that, if you were able to bring the belief about directly on the basis of the demon's threat, you would not be

---

the norm of belief should be taken to be constitutive of the concept of belief, although if my arguments succeed then they would have to find a different argument for this position, other than the fact the claim that the possession conditions for belief involve accepting a normative claim. Pamela Hieronymi [2005] 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.

<sup>16</sup>The *ceteris paribus* clause here is important. We can be rationally criticizable 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.

rationally criticizable for so doing, and thus that the demon's threat would be a reason to believe that the earth is flat after all.

Cases of "motivated irrational" might seem to provide counter examples to this argument. Parfit [1984], 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 force you to do so. 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 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" Parfit [1984, 13]. 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>17</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 (Cf. Parfit [1984];Gauthier [1994]). In these situation it seems perfectly rational to make oneself irrational. However, performing the irrational actions directly would still be rationally criticizable. I would be rationally criticizable 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 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.

---

<sup>17</sup>Due to the amount of harm carrying out my threat will lead to both for myself and for the threatened individual

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) criticizable<sup>18</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 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 being an excellent reason. The same goes for the case of belief.

So it does not seem plausible that it is irrational to do directly what would be rational to bring about indirectly. We can further support this claim by appealing to Andrew Reisner's "argument from blocked ascent". Reisner points out that "you can be offered a deal that will specify that you will lose the prize if you cause yourself to believe the proposition in question, or if you have any higher-order attitudes (in the case of those who argue that there is reason to desire to believe something) towards the first order belief. Ascent can be blocked as part of the pragmatic reason for belief. In the cases

---

<sup>18</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 favor 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 criticizable

in which ascent is blocked, the evidentialist has to bite a very big bullet and accept that there is nothing that counts in favor of holding an evidentially dodgy belief that will prevent the end of the world or make you fabulously rich” Reisner [2009, 271]. In other words, we can imagine an evil demon who not only threatens to destroy the world if you fail to believe the earth is flat, but adds that he will also destroy the world if you take actions to cause yourself to believe this, or form the desire to believe it. In this case, as Reisner points out, the evidentialist would have to hold that we have no reason to prevent the world being destroyed. Furthermore, we can add that they would have to claim that if the person in question did manage to form the belief directly, without resorting to self-manipulation, they would thereby be rationally criticizable. This seems like a highly implausible conclusion. Of course, we might object that forming the belief directly is not criticizable but impossible. I think this is exactly right, and so I accept that as a matter of fact we would have no reason to do as the demon commands, since he is commanding the impossible. 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.

One could object at this point that there is at least some sense in which the person in this example would be criticizable; they would be epistemically irrational. 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. Imagine an agent who could deliberatively form beliefs on the basis of either epistemic or practical 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 favor believing that the world is flat, and the epistemic reasons favor believe 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>19</sup>. Of course, we might 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

---

<sup>19</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. Foot [1972];Copp [2007]). 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. Darwall [1997], Street [2008]).

so, this just shows that the true explanation of transparency lies with the explanation of why positing such a believer is incoherent. It is not the normative truths about belief, or the normative claims accepted by believers, that does the explanatory work here.

## 6 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 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 the 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 transparency can be explained by 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 normativism to provide such a unified account is a mark against it.

## References

- Jonathan Bennett. Why is belief involuntary? *Analysis*, 50:87–107, 1990.
- David Copp. The ring of Gyges: Overridingness and the unity of reason. *Social Philosophy and Policy*, 14(1):86–106, 2007.
- Jonathan Dancy. *Practical Reality*. Oxford: Clarendon Press, 2000.
- Stephen Darwall. Reasons, motives, and the demands of morality: An introduction. In Railton Darwall, Gibbard, editor, *Moral Discourse and Practice: Some Philosophical Approaches*. Oxford University Press, 1997.
- Pascal Engel. Doxastic correctness. *Proceedings of the Aristotelian Society*, 87, 2013a.
- Pascal Engel. In defense of normativism about the aim of belief. In Timothy Chan, editor, *The Aim of Belief*. Oxford University Press, 2013b.
- Gareth Evans. *Varieties of Reference*. Oxford University Press, 1982.

- Philippa Foot. Morality as a system of hypothetical imperatives. *The Philosophical Review*, 81(3):305–316, 1972.
- David Gauthier. Assure and threaten. *Ethics*, 104(4):690–721, 1994.
- Alan Gibbard. Truth and correct belief. *Philosophical Issues*, 15:338–351, 2005.
- Pamela Hieronymi. The wrong kind of reason. *The Journal of Philosophy*, 102(9):437–457, 2005.
- Arie W. Kruglanski and Donna M. Webster. Motivated closing of the mind: "seizing" and "freezing". *Psychological Review*, 103(2):263–283, 1996.
- Ziva Kunda. The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498, 1990.
- Derek Parfit. *Reasons and Persons*. Oxford University Press, 1984.
- Joseph Raz. When we are ourselves: The active and the passive. In *Engaging Reason: On the Theory of Value and Action*, pages 5–21. Oxford University Press, 2002.
- Andrew Reisner. The possibility of pragmatic reasons for belief and the wrong kind of reasons problem. *Philosophical studies*, 145(2):257–272, 2009.
- Kieran Setiya. *Reasons Without Rationalism*. Princeton University Press, 2010.
- Kieran Setiya. Epistemic agency: Some doubts. *Philosophical Issues*, 23:179–198, 2013.
- Nishi Shah. How truth governs belief. *The Philosophical Review*, 112(4):447–482, 2003.
- Nishi Shah. A new argument for evidentialism. *The Philosophical Quarterly*, 56(225):481–498, 2006.
- Nishi Shah and J. David Velleman. Doxastic deliberation. *The Philosophical Review*, 114(4):497–534, 2005.
- Michael Smith. *The Moral Problem*. Blackwell Publishing Ltd., 1994.
- Asbjørn Steglich-Petersen. No norm needed: On the aim of belief. *The Philosophical Quarterly*, 56(225):499–516, 2006.
- Sharon Street. Reply to copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues*, 18:207–228, 2008.
- Sergio Tenenbaum. Knowing the good and knowing what one is doing. *Canadian Journal of Philosophy*, 35:91–117, 2012.
- David Velleman. The possibility of practical reason. *Ethics*, 106:694–726, 1996.
- Ralph Wedgwood. The aim of belief. *Philosophical Perspectives*, 16, 2002.
- Bernard Williams. Deciding to believe. In Howard E. Kiefer and Milton K. Munitz, editors, *Language, Belief, and Metaphysics*, pages 95–111. SUNY Press, 1970.

Bernard Williams. Internal and external reasons. In *Moral Luck*, pages 101–113. Cambridge University Press, 1981.

Nick Zangwill. Directions of fit and normative functionalism. *Philosophical studies*, 91(2): 173–203, 1998.

Nick Zangwill. The normativity of the mental. *Philosophical Explorations*, 8:1–19, 2005.