

| Why Should I Care? Reasons Internalism & Moral Realism: A Reply to Shafer-Landau

Bianca Verjee

Keywords: Reasons internalism, reasons externalism, moral realism, ex post validation, motivation, rational deliberation

Mots-clés: Internalisme des raisons, externalisme des raisons, réalisme moral, validation ex post, motivation, réflexion rationnelle

According to reasons internalism, an agent has a normative reason to act if and only if there is something in the agent's subjective motivational set (their desires, preferences, interests, etc.), or its rational extension, that will be served by so acting. However, for the reasons externalist, some reasons apply to everyone regardless of their particular commitments. Russ Shafer-Landau (2003/2007) is a proponent of moral realism, the view that some moral judgements are objectively true. Reasons internalism serves as a premise in an argument against moral realism. In an attempt to defend moral realism against this argument, Shafer-Landau offers two anti-internalist arguments. This paper considers and rejects both arguments. His first anti-internalist argument is a counterexample designed to show that the internalist restriction on normative reasons is "illegitimate" (Shafer-Landau, 318). I work through several possible avenues for rejecting this argument. My first two arguments show that Shafer-Landau's example is not a counterexample to internalism, as it can be accommodated under the internalist view. My third argument demonstrates that there cannot be a case like the one Shafer-Landau is attempting to construct. I then briefly address Shafer-Landau's second anti-internalist argument, which attempts to show that our moral practices regarding blame and punishment seem incompatible with reasons internalism. For each of my arguments, I consider and respond to some possible objections, and conclude that Shafer-Landau's arguments are not sufficient to warrant rejecting internalism. Thus, his argument in defence of moral realism is weakened.

Selon l'internalisme des raisons, un.e agent.e a seulement une raison normative d'agir si un élément, dans l'ensemble de ses motivations subjectives (ses désirs, ses préférences, ses intérêts, etc.) ou son extension rationnelle, sera servi par cette action. Cependant, pour l'externaliste des raisons, certaines raisons s'appliquent à tous, quels que soient leurs engagements particuliers. Russ Shafer-Landau (2003/2007) est un partisan du

réalisme moral, l'idée que certains jugements moraux sont objectivement vrais. L'internalisme des raisons sert d'une prémisse à un argument s'opposant au réalisme moral. Dans une tentative de défendre le réalisme moral contre cet argument, Shafer-Landau propose deux arguments anti-internalistes. Le présent article considère et rejette les deux arguments. Son premier argument anti-internaliste est un contre-exemple destiné à montrer que la restriction internaliste sur les raisons normatives est « illégitime » (Shafer-Landau, 318). J'explore plusieurs pistes possibles afin de rejeter cet argument. Mes deux premiers raisonnements démontrent que l'exemple de Shafer-Landau n'est pas un contre-exemple à l'internalisme, car son exemple peut être accepté dans le cadre de la vision internaliste. Mon troisième argument démontre qu'il ne peut y avoir un cas comme ce que Shafer-Landau tente de construire. J'aborde ensuite brièvement le deuxième contre-argument de Shafer-Landau, qui essaie de montrer que nos pratiques morales concernant le blâme et la punition semblent incompatibles avec l'internalisme des raisons. Pour chacun de mes arguments, je considère et réponds à certaines objections possibles, et je conclus que les arguments proposés par Shafer-Landau ne suffisent pas pour justifier le rejet de l'internalisme. Par conséquent, son argument en faveur du réalisme moral est affaibli.

Introduction

When do we have a reason to do something? Do moral requirements give us good reasons to act? There are two sorts of answers to these questions. Some assert that the only reasons we have come from our own commitments—our desires, preferences, goals, interests, etc. This position is known as reasons internalism (hereafter, simply internalism)—you have a normative reason to do something because doing it will help you achieve what matters to you. In particular, internalism is the view according to which an agent has a normative reason to act if and only if there is something in the agent's subjective motivational set (their desires, preferences, interests, etc.), or its rational extension, that will be served by so acting. However, especially when it comes to moral reasons—reasons to do what morality demands of us—it seems that some reasons apply to everyone regardless of their particular commitments. This latter position is known as reasons externalism (hereafter, simply externalism)—sometimes we have a reason to do something regardless of our personal desires, preferences, goals, interests, etc. For the externalist, sometimes an agent has a reason to do something regardless of anything internal to that agent, and thus, a reason claim will not be falsified by the absence of a relevant motive (Williams, 292).

Moral realism is the view that some moral judgements are objectively true. Internalism serves as a premise in an argument against this view. In an effort to defend moral realism,

Shafer-Landau offers two anti-internalist arguments. The goal of this paper is to challenge these two arguments.

I will begin with a description of the anti-realist argument and Shafer-Landau's aim, and discuss the features of internalism. I will then present Shafer-Landau's first anti-internalist argument, provide three counter arguments, and address some objections. Finally, I will present Shafer-Landau's second anti-internalist argument and provide a possible route for objecting to this argument.

Shafer-Landau & moral realism

Russ Shafer-Landau (2003/2007) is a proponent of moral realism. According to Shafer-Landau, moral realism "insists on fixing the content of moral demands in a stance-independent way" (312). Internalism serves as a premise in an anti-realist argument, which Shafer-Landau calls the Desire-Dependence Argument:

1. Necessarily, if S is morally obligated to ϕ at t , then S has a good reason to ϕ at t (*Moral Rationalism*);
2. Necessarily, if S has a good reason to ϕ at t , then S can be motivated to ϕ at t (*Reasons Internalism*);
3. Necessarily, if S can be motivated to ϕ at t , then S must, at t , either desire to ϕ , or desire to ψ , and believe that by ϕ -ing S will ψ (*Motivational Humeanism*), and;
4. Therefore, necessarily, if S is morally obligated to ϕ at t , then S must, at t , either desire to ϕ , or desire to ψ , and believe that by ϕ -ing S will ψ .¹

(Shafer-Landau, 312)

If this argument is sound, then the content of moral obligations crucially depends on the agent's commitments. This is incompatible with moral realism. For Shafer-Landau, the conclusion of the Desire-Dependence Argument must be false, for it "tells us, in effect, that any putative moral requirement that fails, or is believed to fail, to fulfil our desires is too demanding, and so *cannot be* morally obligatory" (312). For Shafer-Landau and many others, this goes against our common sense ideas about morality. Since the Desire-Dependence Argument is valid, moral realists must show that the argument is unsound, by rejecting at least one of the premises. Shafer-Landau states that there is no consensus on which premise to

¹ S stands for some subject (or agent), ϕ and ψ each stand for some action, and t stands for some time.

abandon, but attempts to undermine premise 2—reasons internalism. He aims to defend the position that necessarily, there is always good reason to do what morality requires, even if these reasons cannot motivate the agent to whom it applies (Shafer-Landau, 313),

What is internalism?

According to the internalist, the only reasons we have come from our own commitments (our beliefs, desires, long-range projects, and so on) because all reasons must be linked to considerations that are capable of motivating us. Something can be a reason for a person to act only if it presently motivates her, or would motivate her, were she to deliberate soundly from her existing motivations (Shafer-Landau & Cuneo, 283). If this does not happen, the agent has no reason to act. The reasoning behind this view is that “the reasons why an action is right and the reasons why you do it are the same” (Korsgaard, 302). In other words, “[t]he reason that the action is right is both the reason and the motive for doing it” (Korsgaard, 302). As we will see, for Williams, this reason is that the action would serve some consideration in the agent’s subjective motivational set. A reason must imply the existence of a motive, because without a motive, the reason cannot be used to explain the agent’s action. As Korsgaard states, if a reason does not imply a motive, “we cannot say that the person *P* did the action *A* because of reason *R*; for *R* does not provide *P* with a motive for doing *A*, and *that* is what we need to explain *P*’s doing *A*: a motive” (302). If a reason claim did not imply a motive, someone presented with a reason for action could ask why they should do what they have reason to do. Thus, “unless reasons are motives, they cannot prompt or explain actions” (Korsgaard, 302).

In “Internal and External Reasons,” Williams asserts that an agent has a reason to ϕ if and only if there is something in the agent’s *subjective motivational set*—their “*S*” (292)—that would be served, or that the agent *believes* would be served, by ϕ -ing. In addition to desires, “*S* can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects ... embodying [the agent’s] commitments” (Williams, 294). It is important to note that not all elements in an agent’s *S* will necessarily be egoistic (Williams, 294); one can, for instance, have a desire to help homeless individuals. A reason statement must be relative to an agent’s *S* because, in order to explain their action, the reason we cite must be capable of having motivated them to so act (Williams, 293).

An agent can add to their *S* through rational deliberation. For instance, rational deliberation might lead to the conclusion that one has reason to ϕ because ϕ -ing would be the most convenient, pleasant, or economical way to satisfy some element in one’s *S*.² This means one’s reason to ϕ can come from their existing motivations, or from additional motivations they

² While this is an especially clear example of practical reasoning leading to conclusions about what one has reason to do, Williams notes that there are “much wider possibilities for deliberation” (294).

would have after rational deliberation. Among other things, deliberation can involve imagination and persuasion from others. As a result of these processes one can come to see that she has a reason to do something, which she had not realized she had reason to do. The deliberative process can also subtract elements from *S*. For instance, it can lead an agent to discover that some belief is false, and thus come to realize that she actually has no reason to do something she thought she had reason to do. Since the deliberative process can add new actions for which we have internal reasons, add new internal reasons for actions, and subtract elements from our *S*, we should not think of *S* as static. (Williams, 294)

We can now see why internalism serves as a premise against moral realism. Internalism is incompatible with moral realism because moral realism requires that at least some moral facts or values be objective. That is, some moral facts or values are reason-providing for every individual, regardless of what they care about. However, if internalism is true, there are no such reasons; whether a moral fact or value provides an individual with a reason to ϕ depends on whether there is something in the individual's *S* that would be served by ϕ -ing.

It is important to note that internalism requires that the member (or members) of *S* that would be served by ϕ -ing "succeed in motivating us [only] insofar as we are rational" (Korsgaard, 305). In other words, *R* is a reason for agent *A* to ϕ only if *A* would be motivated to ϕ if *A* were rational. Thus, internalist reason statements are meant to apply to agents only insofar as they are rational. Williams seems to agree, stating that "the internal reasons conception is concerned with the agent's rationality" (293). Since agents are not always rational, they will not always be aware of the reasons they have. This also means they may not currently have elements of *S* that they would have if they were to rationally deliberate.

Korsgaard states that "to act irrationally [...is...] to fail to be motivationally responsive to the rational considerations available to us" (304). This failure could, for instance, be the result of "some physical or psychological condition" (Korsgaard, 304). However, the agent still has reason to act, "for all that is necessary for the reason claim to be internal is that we can say that, if a person did know and *if nothing were interfering with her rationality*, she would respond accordingly" (Korsgaard, 304).

Shafer-Landau's first argument against internalism

Shafer-Landau's first anti-internalist argument is a counterexample to internalism, intended to show that this internalist restriction on normative reasons to only those linked to the agent's *S* or its rational extension, is illegitimate. He asks the reader to consider someone who expects the worst due to her pessimism, negative self-image, shyness, and reluctance to take risks, and to suppose that she would in fact gain pleasure "were she to emerge from her shell" (Shafer-Landau, 318). In fact, "the value of [these] experiences would have been endorsed

by the agent herself, after she has had the benefit of those experiences" (Shafer-Landau, 319). "It is this *ex post* validation that makes it true to say of her, in her earlier phase, that she had a reason to extend herself" (Shafer-Landau, 318). However, because of her attitude and personality, she does not see that she would benefit from these actions, despite her having good reason to do so. "[N]othing in her existing motivations [her *S*] would lead her to take [these] steps" (Shafer-Landau, 318). Since internal reasons statements are "falsified by the absence of some appropriate element from *S*" (Williams, 293), internalists are committed to saying that she has no reason to engage in these beneficial activities, which seems quite counter-intuitive. This suggests that internalism is the wrong view.

For the sake of simplicity, I will call this agent Clara and will suppose that the action she would gain pleasure from is going to a party.

The first avenue to rejecting this argument I will explore, is to show that Shafer-Landau's example is not a counterexample to internalism because his example suggests that there is at least one element in the agent's *S* that is preventing her from seeing that she has reason to go to the party, and that element is grounded in a false belief.

Either Clara has nothing in her *S* that would motivate her to go to the party, or, there is something in her *S* that motivates her *not* to go to the party. Shafer-Landau is not completely clear on which of these situations his example is supposed to emulate. However, he describes Clara as having an aversion to the idea of going to the party, which suggests the latter of the two possible interpretations.

According to Shafer-Landau, Clara anticipates no pleasure from going to the party, and the dread she experiences when imagining it, prevents her from taking any steps towards going to the party (318). This, combined with the fact that going to the party would in fact bring her pleasure, suggests that something in her *S* is preventing her from seeing that she would benefit. Since she would in fact benefit, that element in her *S* must be grounded in a false belief. For instance, she may believe that in order to go to a party, you have to be very extroverted, confident, and interesting. Since she considers herself to be more of an introvert, is quite shy, and considers herself quite uninteresting, she anticipates embarrassment and awkwardness, and thus has a desire to avoid parties. This desire to avoid parties, which is an element of her *S*, is grounded in false beliefs; she is not uninteresting, and it is not true that one needs to be extroverted and confident in order to go to a party.

However, according to the internalist, elements in *S* that are based on, or grounded in, false beliefs cannot give rise to legitimate internal reasons. Suppose, Considering Williams' gin and tonic example, you desire a gin and tonic. You believe the liquid in front of you is gin, but it is in fact petrol. Do you have reason to mix the liquid in front of you with tonic and drink it (Williams, 293)? Most people's intuition, including my own, is that you do not have reason to mix

the liquid with tonic and drink it. In fact, you have every reason *not* to drink it. The item in your *S* grounded in false belief in this scenario is the desire to drink what is in front of you. You want to drink it because you believe it is gin, but this belief is false. Thus, you *think* you have reason to do something, but in fact, you do not have reason to do it. Likewise, Clara has a desire to avoid parties, but, given that she would actually really enjoy the experience, she must have a false belief which is grounding this desire. The consequence of this is that she thinks she has reason to avoid parties, but in fact, she has no such reason. In other words, she thinks she has no reason to go to the party, when in fact, she does have reason to go to the party.

Williams states that an agent may falsely believe an internal reason statement about himself (293). Based on the above discussion, it seems that Clara falsely believes the following reason statement about herself: that she has reason to avoid going to the party. Williams also states that an agent may not know some true internal reason statement about themselves, and that one reason for this is that the agent may not know some fact such that if they did know it, they would be disposed to ϕ , in virtue of some element of their *S* (293). Based on what was discussed above, it seems that Clara does not know the following true reason statement about herself: that she has reason to go to the party; and that furthermore, upon discovering the fact that her beliefs about the party are false, she would be disposed to go to the party, in virtue of some element in her *S*. Thus, this is not a counterexample to internalism; the reason that she is unable to get from her current state to the conclusion that she has reason to go to the party by rational deliberation is not that there is no such sound deliberative route, but rather that this route is blocked by the element of her *S* that is grounded in a false belief. The internalist can say here that the agent *would* see that she has reason to go to the party were it not for the item in her *S* that is grounded in a false belief.

One might object here by arguing that not all false predictions about whether one would enjoy an experience, are grounded in false beliefs. However, I find it hard to conceive of a case in which one predicts that an experience will be unpleasant, finds that it is actually enjoyable upon trying it, but where all their beliefs relevant to that prediction are true. This is because, for a rational person, their predictions are grounded in their beliefs. Consider an individual who is averse to some experience and predicts no pleasure from it, but, upon trying it out, learns that they actually like the experience. What did the individual learn if not that some of her beliefs relevant to what the experience will be like, were wrong? In such a case, it seems that, by trying the experience, the individual has learned that their prior outlook was not completely accurate. Since one's prior outlook is presumably comprised of beliefs about the experience in question, there must have been some such beliefs that turned out to be false. Therefore, the agent's aversion when imagining what the recommended experience will be like, combined with the fact that she would actually enjoy the experience, suggests that there is an element in her *S* grounded in false belief, that is preventing her from seeing that she has reason to go to the party.

However, there is one problem with this avenue to rejecting Shafer-Landau's argument. Shafer-Landau cannot have meant his example to be interpreted in this way. He must have intended that there be no element in her *S* from which a sound deliberative route could lead to the conclusion that she has reason to go to the party, rather than that there be an element of her *S* blocking that route, as this is what would be necessary to produce a counterexample to internalism. Thus, in order to be charitable, we must assume he meant the former—that there is nothing in Clara's *S* that would lead her to conclude that she has reason to go to the party.

The second avenue to rejecting Shafer-Landau's counterexample I will explore is to show that, though it is possible that such an agent could exist, that agent would not be rational; if she were rational, then she would come to know that she would find pleasure in going to the party, and thus see that she has reason to go to the party. Thus, Shafer-Landau's example can be accommodated under internalism and is therefore not a counterexample to internalism.

Shafer-Landau describes several important features of his example. The first is that Clara has certain attributes, including melancholia, pessimism about future happiness, shyness, and poor self-conception (Shafer-Landau, 318). Here the internalist can argue that these psychological conditions prevent Clara from deliberating rationally. For instance, upon imagining what it would be like to go to the party, Clara might only be imagining the things that could go wrong, due to her pessimism. However, in considering a choice, a rational person considers both the positives and the negatives. Without this interference from her psychological conditions, she would be properly motivated by these reasons.

The second feature of Shafer-Landau's example is that Clara is "clear-headed enough, and can imagine the experiences of mingling, social chat, and light flirtation" (Shafer-Landau, 318). But this does not guarantee that Clara is fully rational, for it is possible to understand the argument theoretically, without understanding the practical implications. According to Korsgaard, "[a] person in whom [the motivational path from ends to means] is, for some cause, blocked or nonfunctioning may not respond to argument, even if this person understands the argument in a theoretical way" (306). Since this possibility is not ruled out by anything in the example, it remains a possible objection to those that would claim the agent is rational. However, it is important to note that this point on its own would not be sufficient to establish that Clara is irrational. For though it is possible that she does not understand the practical implications, the most charitable way of reading Shafer-Landau's claim is that Clara understands both the theoretical argument, and its practical implications.

Third, Clara anticipates no pleasure from going to the party, and the dread experienced when she imagines the experience prevents her from taking any steps towards going to the party (Shafer-Landau, 318). This, combined with the fact that these experiences would in fact bring Clara happiness, gives one reason to think that Clara is not rationally considering the possible and likely outcomes. For, at least one possible outcome of going to the party is a

positive one; a rational agent would be able to see these possible positive outcomes, in addition to the negative ones. There are also likely several things Clara would like about the party; surely a rational agent would be able to see at least some of these.

Fourth, “nothing in her existing motivations would lead her to take [the recommended] steps” (Shafer-Landau, 318). But this claim does not consider the possible rational extensions of Clara’s *S*, in which case it may be that she would have an element in *S* that would motivate her to act, were she to rationally deliberate. Furthermore, even if we take Shafer-Landau to be referring to both existing motivations and their rational extensions, it could be that she is unable to see that she has reason to act, due to her present (irrational) state.

Earlier in his paper, Shafer-Landau states that, for internalists, “[t]here can be a sound deliberative route from one’s motivations to a reason even if psychological impediments prevent one from ever being able to trace such a route” (313-314). This gives even more support to the possibility that the agent has reason to act, despite her being unable to see this in her current state.

Shafer-Landau takes his example to demonstrate several things. First, “[i]t is true of many ... that if they were somehow to ‘look beyond’ the picture of things they have grown used to, they would find themselves with an outlook, a plan of life, and set of circumstances that they would find more valuable than they could ever have imagined” (Shafer-Landau, 318). However, the phrase ‘look beyond’ seems to describe thinking more rationally—taking a step back from one’s current feelings and looking at things more objectively. Since internalism requires only that the agent be motivated insofar as she is rational, and the above point seems to suggest that she would be so motivated, this provides no issue for internalism.

Second, “realizing the relevant benefits often requires a change of character” (Shafer-Landau, 318). While this seems a large barrier to realizing the benefits in question, the internalist could argue that the only change required is a more rational outlook. Nothing in the agent’s core nature need change. In order to see that we have reason to do something, we need not change our goals, personality, commitments, etc. We merely need to consider the possible benefits of what we supposedly have reason to do, from a more rational perspective.

Third, “prior to this change [in character], the prospects of the new life do not appeal, just because they are rationally unrelated to one’s present outlook” (Shafer-Landau, 318). However, although the prospects of the new life may be rationally unrelated to one’s present outlook, because it is an irrational one, this does not mean they are rationally unrelated to one’s present *situation*. Clara’s present outlook—her attitude, mood, and perspective—is a pessimistic, and thus, irrational one. However, her situation is that she has certain desires, interests, likes, etc., and doesn’t have many friends. One’s outlook changes more easily and frequently, while one’s situation is more stable. For instance, I may not feel like starting my

paper because I am so frustrated with the class in which the paper is due and am struggling to understand the material. However, I may still have a reason to start my paper (though I do not realize it at the moment) because I desire to do well in the course and need to start as soon as possible so that I have ample time to complete it. In this scenario, my current outlook—my frustration—is preventing me from thinking rationally. However, were I to think more rationally, I would see that I have reason to start my paper now. Likewise, although Clara may not see that she has reason to go to the party, due to her current outlook, she may still benefit from going to the party given her desires, interests, likes, and the fact that she might make some friends. She may simply be unable to see this relation due to their present (irrational) outlook. All that is required is that she would be motivated to act, if she were rational.

Fourth, the recommendations will likely “fall on deaf ears” (Shafer-Landau, 319). This is unproblematic because not being convinced by good reasons is compatible with nevertheless having good reasons, so long as it is true that the agent is not rational. According to Korsgaard, “it will not always be possible to argue someone into rational behaviour” (306). However, this does not mean they have no internal reason to act. It only means they are not rational enough to see it. Since I have already argued for Clara’s irrationality, her not being appropriately motivated by the available reasons is not an issue.

From all this, Shafer-Landau concludes that his example describes a situation in which the agent has reason to act, despite there being nothing in her *S* that would motivate such actions, and so internalism must be false. However, having provided internalist objections to each of Shafer-Landau’s points, it seems that we have shown that his example can be accommodated by internalism. Clara’s failure to see the benefits of acting can be just as easily explained by claiming she is irrational, as by claiming she ought to be motivated by things unrelated to her *S*. If this is true, then Shafer-Landau’s argument fails to show the illegitimacy of the internalist restriction on normative reasons.

One might object here by pointing out that my proposed internalist response, and indeed Korsgaard’s view of internalism, “falls short of a defense of internalism” (Shafer-Landau, 315) because it only guarantees a “motivational link” (Shafer-Landau, 315) between the agent and their reasons when the agent is practically rational. According to Shafer-Landau, a defence of internalism needs to show that reasons are “capable of motivating us full stop” (315). However, Korsgaard has only shown that “our reasons must be capable of motivating us to the extent that we are practically rational” (Shafer-Landau, 315). Since we are not always practically rational, her view does not amount to a defence of internalism. If Shafer-Landau is right, then my objections lose much of their force.

This objection takes it to be problematic that reasons are only capable of motivating us insofar as we are rational. But how could it be otherwise? I do not think it should be a problem for the internalist that an irrational individual may not be properly motivated by their reasons.

For, to be rational is to respond appropriately to our situation. When we are irrational, our faculty of reasoning is not working as it should be. Thus, the agent “fails to make the rational connection” (Shafer-Landau, 315) between their subjective motivational set and the recommended actions, or succeeds in making the connection, but “fails thereby to be motivated” (Shafer-Landau, 315), possibly owing to “psychological infirmities” (Shafer-Landau, 315).

For someone whose rational faculties are not properly functioning, we cannot say anything of what we should expect them to do. Irrationality is unpredictable—it means the agent is not acting or thinking as they should. For someone who is in such a state, we cannot say anything of how they will think or act. Such agents may respond appropriately to their reasons, but they may not. Thus, the internalist cannot be expected to provide an account of how motivation, thought, and action work in an irrational individual because the fact that they are irrational means that something is just not *working* in the first place.

In fact, that an agent is only motivated by their reasons if they are rational, seems to be a benefit of internalism, for it explains why some individuals do not act on their good reasons. On the internalist view, a rational individual and an irrational individual may both have good reason to ϕ , but it may be that only the rational individual is motivated by this reason. This explains why many people do not act in accordance with their reasons. It seems to me that a theory of moral reasons divorced from the particulars of the agent and current situation will have a more difficult time explaining why some individuals are motivated by their good reasons, while others are not. Thus, the internalist’s theory need not explain the agent’s behaviour when that agent is irrational. It is sufficient to provide an internalist account of moral reasons that guarantees a “motivational link” (Shafer-Landau, 315) only for those who are rational. Thus, this objection does not present a problem for my argument that Shafer-Landau’s example can be accommodated under internalism.

While both these strategies (arguing that there is an element in Clara’s *S* grounded in false belief, and showing how each feature of the example can be accommodated under internalism) may respond to the example as it is presented by Shafer-Landau, it does not address what Shafer-Landau likely intended his example to do. In particular, there are two key features that Shafer-Landau likely intended his example to have. The first is that there is no element in Clara’s *S* or its rational extension from which to draw the conclusion that she has reason to go to the party. We see evidence of this intention in Shafer-Landau’s statement that his example “makes its point only if the appeal does not contain a rational relation to the addressee’s existing motivations” (Shafer-Landau, 319). The second is the idea of *ex post* validation; were Clara to go to the party, she would in fact find pleasure and value in it. It is these two key features that produce a counterexample to internalism—a case in which an agent has reason to ϕ , despite there being nothing in her *S* or its rational extension that would

motivate her to do so. I will discuss both of these intended features, and argue that such a case could not exist. I will then consider some objections to my argument.

In my first argument, I argued that there was some element in Clara's *S* that was preventing her from seeing the sound deliberative route from her current position, to the conclusion that she has reason to go to the party. In my second argument, I argued that though an agent such as Clara could exist, she would not be rational. However, the externalist can reply to these claims by pointing out that in Shafer-Landau's example, it is supposed to be true, *ex hypothesi*, that nothing in Clara's *S* or its rational extension would lead her to go to the party. In other words, there is supposed to be no item in Clara's *S* or its rational extension from which there is a sound deliberative route to her being motivated to go to the party. This is the first key feature of the counterexample Shafer-Landau intended; it is supposed to be the case that no amount of rationally accessible information will get Clara to see that she has reason to go to the party, because there is no element in her current *S* or its rational extension from which to reason from in order to arrive at the conclusion that she has reason to go to the party. These are meant to be constraints on the case, and thus, the internalist responses I have provided so far have been unsuccessful.

The second key aspect of Shafer-Landau's intended counterexample is the idea of *ex post* validation. He asks the reader to suppose that, were Clara to "emerge from her shell" (Shafer-Landau, 318), or in our case, attend the party, "she would find new pleasure, even some delights" (Shafer-Landau, 318). He says that she would consider herself much better off than if she had not gone to the party, and would come to endorse the value of the experience from within. "It is this *ex post* validation," says Shafer-Landau, "that makes it true to say of her, in her earlier phase, that she had a reason to [attend the party]" (318). Together, these features create a case in which the agent has reason to do something, even though nothing in her *S* would motivate her to do it.

This seems, at first, to be a plausible scenario. Sometimes you do not know that you will like something until you try it, and thus, you cannot be rationally expected to know that you would like it without trying it. To illustrate this idea, we can consider Frank Jackson's example of Mary the colour scientist. While the example was not intended for this purpose, it can be used to help illustrate the idea that we can learn new things from experience that no amount of information or rational deliberation will help us to discover, and to reveal an important dimension of Shafer-Landau's intended counterexample.

According to Jackson, Mary is an extremely bright scientist who has been forced to learn about the world from a black and white room through a black and white television screen. She specializes in the neurophysiology of vision and has acquired all the physical information there is about colour and what goes on when we see colour. Jackson asks us to consider what will happen when Mary is freed from the black and white room, or is provided with a colour

television screen. He states that “[i]t just seems obvious that she will learn something about the world and our visual experience of it” (Jackson, 130). In other words, Mary will learn from her experience something she never could have learned, no matter how rational she was and how much information she had. Likewise, it may be the case that Clara will learn something about parties that she could never have learned from more information or rational deliberation. Both Mary and Clara gain something new from their experience.

However, there is an important difference between Jackson and Shafer-Landau’s examples. While Jackson’s example is meant to demonstrate the existence of *qualia* (subjective phenomenal features that accompany experience), Shafer-Landau’s example is meant to demonstrate that Clara will enjoy and find value in the experience. For Clara, the point is not merely that, by going to the party, she will know what it feels like to go to a party. Rather, Shafer-Landau’s point is that, from these feelings, she will conclude that the experience is enjoyable and valuable to her. Having highlighted a key dimension of Clara’s case, my next task will be to illustrate the relevance of Clara drawing a conclusion from her novel experience.

Suppose Clara actually does go to the party, and does in fact find it extremely enjoyable. What explains why she enjoyed it? Surely it must be something about Clara and what she likes and values; presumably, Clara enjoyed the party because it had features which were attractive or valuable to her. Perhaps she liked it because she finds chatting with others fun and interesting, or perhaps she liked getting dressed up. Whatever the explanation, it seems to be the case that any explanation will refer to some feature of the party that corresponds to Clara’s characteristics (her values, likes, interests, etc.). In other words, an explanation of why Clara found the experience enjoyable must refer to some element or elements in her *S*. If this is the case, then it seems that it should have been possible for Clara to rationally deliberate from this element, call it S_1 , to the conclusion that she would enjoy the party and thus, has reason to attend it (assuming that she is fully rational, and has epistemic access to the elements of her *S*). Now she may not take the time to fully reason this out and rationally deliberate, and thus, may not actually discover S_1 , and come to the conclusion that she has reason to go to the party. However, so long as it is possible for her to do so, she can be considered as having an internal reason to go to the party.

In short, I do not doubt that there are experiences out there for which we will never know what it feels like until we try it. What I do doubt is that one could have such an experience, enjoy the experience, *and* be able to explain why it was enjoyable without referring to anything in their *S* or its rational extension. There is a tension between the lack of a relevant element in *S*, and the *ex post* validation, that I cannot reconcile. Thus, I do not think such a case is possible.

One could object here by stating that perhaps the party added a new item to her *S*. Williams does note that an agent’s *S* should not be considered as statically given—items can be

both added to and removed from it. So, perhaps nothing in Clara's *current S* would have led to the conclusion that she would enjoy the party, but after the party, a new item is added to her *S*, explaining why she enjoyed it. For instance, perhaps Clara does not know that she likes getting dressed up until she tries it. For the internalist, your reasons depend on where you are now; perhaps Shafer-Landau is trying to exploit this. He does note that one may have reason to change their present outlook (Shafer-Landau, 318). He states that "the goods available only to those who make such changes may be so valuable as to make it true that one has, despite one's present motivations, a reason to make the necessary changes" (Shafer-Landau, 318-319).

Shafer-Landau seems to be saying here that one may come to see the value in some activity if they were to change their outlook, traits, preferences, motivations, or some other element, or elements, in their *S*. While I recognize that people's personality and attributes change over time, I find it hard to conceive of what could be meant by something like: "You would have enjoyed it if you'd tried it" except that, based on the sort of person you are *now*, your characteristics, your mental states, etc., you would have enjoyed the party, had you attended it. When you say to someone: "You would enjoy it," you are presumably saying it because you believe it will be valuable to *that person* in particular, which means that there is at least one thing about the person as they are now (their values, personality traits, interests, goals, etc.) that makes you think they would enjoy it. This fact about the person that grounds your belief or assertion that they will enjoy the activity, would be a member of their current *S* (or its rational extension). Thus, it seems it cannot be the case that Clara would enjoy the activity, but nothing in her present situation would lead to that conclusion.

In sum, I have argued that, even when we take a more charitable interpretation of Shafer-Landau's example, considering the counterexample he likely intended to create, it seems that such an example could not exist. Thus, Shafer-Landau's counterexample argument does not succeed in undermining internalism.

Shafer-Landau's second argument against internalism

Up to this point I have only addressed the first of Shafer-Landau's two anti-internalist arguments. I would now like to leave the reader with a consideration that may cast some doubt on his second argument as well.

Shafer-Landau asks us to consider a person who has a strong dislike of others, is completely indifferent to what others think, and is so determined to be cruel, that nothing in his *S* or its rational extension would prevent him from committing the most heinous crimes, and so, it is irrational for him to refrain. Shafer-Landau states that such a scenario is a problem for the internalist because we tend to think that people have reason to refrain from such behaviour, regardless of their personal commitments—of what they care about. However, if

internalism is true, then since there is no rational deliberative route from some element in their *S* to the conclusion that they have reason to avoid committing these heinous crimes, the person in question has no reason to refrain from such crimes. But since blame requires “failure to adhere to good reasons” (Shafer-Landau, 319), and such people have no reasons to avoid these evil deeds, they are morally blameless. Such people also presumably cannot be punished, since punishment is “predicated on blameworthiness” (Shafer-Landau, 319). Thus, if internalism is true, such people cannot be justly blamed or punished for failing to refrain from these actions. Since such people are, in fact, justly blamed and punished, internalism must be false (Shafer-Landau 319).

There are two questions at issue here. The first is whether one can blame an agent for doing something which, on the internalist view, they had no reason to avoid. The second is whether one can rightly punish such a person. I interpret blame as consisting of attitudes toward the agent in question, including our judgements about what they should have done, and what kind of person they are in virtue of not doing what they ought to have done. In contrast, I consider punishment to be the actions we take towards them to deter future crimes, such as limiting their freedom by putting them in jail. I am willing to bite the bullet that we cannot justifiably blame an agent who could not be reasonably expected to have acted otherwise. However, it is possible that we could be justified in punishing such agents. The agents we are considering act because there is nothing in their *S* or its rational extension to motivate them to refrain. So, it seems that what is needed is a reason for them not to perform such actions in the future. Punishment can provide such a reason; it can add to their *S* a desire to avoid punishment. This also means they can be rightly blamed *and* punished for failing to avoid these actions in the future.

This certainly requires an explanation of why such agents ought to have reason to refrain from these actions—an issue that is beyond the purview of this paper. However, it at least leaves open the possibility of a route to objecting to this second of Shafer-Landau’s anti-internalist arguments.

Implications

Up The ideas developed in this paper regarding reasons internalism have implications for political science, sociology, anthropology, and psychology. A deeper understanding of reasons internalism can help us to better understand the causes of human behaviour in terms of how individuals respond to their reasons and what motivates them to act. This can help us to better understand and predict human behaviour. For instance, throughout this paper, we have seen that individuals are motivated by items in their subjective motivational set, such as desires, preferences, and interests. As we have seen, what reasons an individual has, and whether they will respond to those reasons, depends on whether they are rational, what

psychological conditions are interfering with their ability to reason, whether they have taken time to rationally deliberate, what facts and information they have taken into account in this deliberation, and whether they have any false beliefs. Considering each of these factors can help us to more accurately determine what a given individual might do.

In addition, my discussion of our moral practices regarding blame and punishment has important implications for law and governance. In particular, my argument highlights the importance of having laws with clearly specified punishments. This is essential because, for those in society who are not otherwise motivated to refrain from undesirable and harmful behaviours, the threat of punishment creates a desire to avoid punishment, which then creates an internal reason to refrain from those harmful behaviours. If citizens are unaware of the punishments for failing to refrain from some harmful behaviour, they may be less inclined to refrain. The government must also ensure that it strongly enforces its laws and is consistent in its dispensing of punishments. For, if some individuals are punished while others are not, citizens who lack other motivations to refrain from crime will have less reason to refrain. Furthermore, the government must ensure that the punishments are severe enough that an individual who is considering whether to commit a crime, will have reasons to refrain that are stronger than their reasons to commit the crime. In other words, their desire to avoid the punishment in question must be able to compete with and trump their motivations for committing the crime.

Conclusion

At the start of this paper, I explained that Shafer-Landau's anti-internalist arguments were a necessary component of his defence of moral realism. However, I have shown that Shafer-Landau's first argument—his counterexample—can, in fact, be accommodated under the internalist view, and even a more charitable interpretation of his example fails, as there cannot be a case like the one he is attempting to construct. I have also cast doubt on Shafer-Landau's second argument by asserting that the threat of punishment can give an agent a reason to avoid committing crimes, despite a lack of other motivations to refrain, thus allowing us to justly blame and punish such agents, in keeping with our current moral practices. Having demonstrated that his first argument fails, and casting doubt on his second, I have shown that Shafer-Landau's arguments are not sufficient to warrant rejecting internalism, and thus his argument in defence of moral realism is weakened.

In addition, I hope that my exploration and analysis of reasons internalism has provided the reader with a deeper understanding of human motivation and action with which to better predict human behaviour and inform our governing practices.

References

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32 (127): 127-136.
<https://doi.org/10.2307/2960077>

Korsgaard, Christine. (1986) 2007. "Skepticism about Practical Reason." In *Foundations of Ethics: An Anthology*, edited by Russ Shafer-Landau and Terence Cuneo, 299-311. Hoboken: Blackwell Publishing. <https://doi.org/10.1111/j.2041-6962.2001.tb01806.x>

Shafer-Landau, R., and Terrence Cuneo. 2007. *Foundations of Ethics: An Anthology*. Hoboken: Blackwell Publishing.

Shafer-Landau, Russ. (2003) 2007. "Moral Reasons." In *Foundations of Ethics: An Anthology*, edited by Russ Shafer Landau and Terence Cuneo, 312-327. Hoboken: Blackwell Publishing.

Williams, Bernard. (1979) 2007. "Internal and External Reasons." *Foundations of Ethics: An Anthology*, edited by Russ Shafer-Landau and Terence Cuneo, 292-298. Hoboken: Blackwell Publishing.