

Internalism without Luminosity¹

Abstract:

Internalists face the following challenge: what is it about an agent's internal states that explains why only these states can play whatever role the internalist thinks these states are playing? Internalists have frequently appealed to a special kind of epistemic access that we have to these states. But such claims have been challenged on both empirical and philosophical grounds. I will argue that internalists needn't appeal to any kind of privileged access claims. Rather, internalist conditions are important because of the way in which we expect them to act as *causal* mediators between states of the world, on the one hand, and our beliefs and actions on the other.

1. Introduction

My goal is to defend a moderate form of internalism without appealing to *luminosity*: the claim that there is a non-trivial set of conditions such that, whenever we are in them, we are in a position to know that we are in them. In a nutshell, the internalist claim that I will be defending is the claim that there are *some* purposes for which internalist epistemic norms are important in epistemology. But to state the thesis precisely, it will be helpful to situate it within the context of the internalism/externalism debate.

Internalism about justification is generally understood as the view that one's justification supervenes on some set of internalist conditions. Justification-internalists come in a variety of stripes² but, for the purposes of this paper, I will understand internalist conditions as conditions that describe an agent's current non-factive mental states. The externalist about justification is just someone that denies justification-internalism. Two externalist theories that will feature prominently in this paper are:

- (1) *Reliabilism*: A belief is justified if and only if it was produced by a reliable process (e.g. Goldman (1979)).

¹ Many thanks to Charity Anderson, Brian Cutter, Sinan Dogramaci, Daniel Greco, Jeff Russell and David Sosa, and especially to John Hawthorne and Susanna Rinard for invaluable discussion and feedback on earlier drafts. I would also like to thank audiences at UT-Austin's Epistemology Reading Group, UNC Chapel Hill, University of Wisconsin at Madison and participants in the Texas Epistemology Extravaganza. Some of the work for this paper was conducted at Oxford University and supported by a grant to the *New Insights and Directions in Religious Epistemology* project funded by the John Templeton Foundation.

² See Pryor (2001) for a survey.

(2) *Knowledge First*: A belief is justified if and only if it constitutes knowledge. (e.g. Williamson (2000)).

Another way of expressing the thought that one's justification should depend only on a set of internalist conditions is to claim that whether one is justified in adopting some belief is determined by whether one conforms to a set of *internalist epistemic norms*. By "*epistemic norm*" I just mean any conditional of the form: "In circumstances C, adopt doxastic state d." An *internalist epistemic norm* is a norm whose antecedent is an internalist condition. So we can also think of justification-internalists as those who think that justification is governed by a set of exclusively internalist norms, while justification-externalists deny this claim.

So who is right? If you don't like picking sides, you can be a pluralist. Indeed, a variety of philosophers have thought that internalist and externalist norms are *both* worthy of our consideration. Here, for example, is Richard Foley:

One project is to investigate what has to be the case in order to have knowledge. An externalist approach is well suited to this project. A distinct project, also important, is concerned with what is required to put one's own intellectual house in order (2001, 21).

Plantinga (1993) thinks that the internalist notion of justification, though not suitable for what he is interested in (warrant) is "a fine thing, a valuable state of affairs (45)." What the internalist notion captures, for Plantinga, is "a deontological matter, being permitted or within one's rights, conforming to one's intellectual obligations" (46). Relatedly, internalist norms have been thought by some to be appropriate for tracking blameworthiness (Steup (1999)), and even Goldman (1988) has acknowledged that internalist norms may be well suited for this task.³

The people who I will call *hardcore externalists* are people who think that internalist norms play *no* interesting role in epistemology. Timothy Williamson (2000) is a paradigmatic hardcore externalist, and Amia Srinivasan's piece, appearing in this volume, makes a strong case for hardcore externalism as well. Additionally, papers by Greco (2005), Goldman (1999), Jacobson (1997), and

³ Others pluralists include Montmarquet (1993), Ernest Sosa (2001), Alston (2005) and David Sosa (forthcoming).

Kornblith (1988) contain arguments that challenge claims concerning the purposes for which internalist norms are, allegedly, especially well suited (for example, deontological purposes, guidance giving, and blame).

My aim is to argue, without appealing to luminosity, against hardcore externalism, and so in favor of moderate internalism: the claim that internalist epistemic norms have *some* important role to play in epistemology. In the next section, I will motivate hardcore externalism and explain why it is, indeed, difficult to be an internalist of any sort without appealing to luminosity. The remainder of the paper is devoted to addressing this challenge. I will argue that internalist norms deserve our attention, not because of any special epistemic access that we have to internalist conditions (though perhaps we have that too), but rather because of a special *causal* role that we should expect internalist conditions to play.

2. Why be a Hardcore Externalist?

Why would one think that there is *no* interesting role for internalist norms to play in epistemology? In this section, I describe one motivation: the idea that a commitment to any special role for internalist norms is ultimately rooted in a commitment to the luminosity of internalist conditions. The motivation I will describe is inspired by the work of Alston (1993), Kornblith (1988), Goldman (1999), Williamson (2000) and Srinivasan (this volume).

Here's the idea: If you're a moderate internalist, you think that internalist norms play *some* important role in epistemology – call it R. The hardcore externalist I have in mind thinks that such an internalist must respond to the following challenge:

The Externalist Challenge: Take whatever role R that you think a set of exclusively internalist norms are especially well suited to play. Explain why a set of externalist norms can't *equally* well play role R.

To make this concrete, suppose you believe that only a set of exclusively internalist epistemic norms can track *epistemic blameworthiness*. What the hardcore externalist wants from the blameworthiness-internalist is some explanation as to why it is only

internalist conditions that can bear on whether an agent is blameworthy. Why can't externalist conditions also bear on whether an agent is blameworthy? The hardcore externalist I have in mind suspects that the answers the internalist will provide in response to this question will rely on a false assumption about a kind of privileged access that we have to internalist conditions. As an illustration, consider appealing to the following claim as a way of responding to the externalist's challenge:

BLAMEWORTHINESS: If N is a norm that determines blameworthiness, then, whenever the antecedent of N obtains, an agent must be in a position to know that it obtains.

The reason that appealing to BLAMEWORTHINESS is tempting is that if you start thinking about why it seems inappropriate to blame the unfortunate brain-in-a-vat for her false beliefs, despite the fact that she has a terribly unreliable perceptual system, one is inclined to say: *because there's no way for her to know that her perceptual system is unreliable*. This suggests that you can't blame someone for failing to comply with a norm in circumstances in which they can't tell whether the antecedent of the norm obtains. If that's right, then a norm governing blameworthiness *itself* must be one whose antecedent is such that *whenever* it obtains, one is in a position to know that it obtains.

Now, to get from BLAMEWORTHINESS to the claim that only internalist epistemic norms can govern blameworthiness we need a principle linking the sorts of conditions that one is always in a position to know about to internalist conditions. So to complete the blameworthiness response we need:

B-LINK: Only internalist conditions are such that, whenever they obtain, agents are in a position to know that they obtain.

On this view, what makes internalist conditions especially well suited to serve as the grounds for blameworthiness is that they are the only kinds of conditions whose obtaining we are always in a position to know about. And this is where the hardcore externalists get off the boat. For, they claim, if there are to be *any* nontrivial norms that the internalist thinks determine blameworthiness, it follows that:

LUMINOSITY: There is a nontrivial set of internalist conditions such that whenever those conditions obtain, an agent is in a position to know that they obtain.

But LUMINOSITY has been challenged on both philosophical and empirical grounds.⁴

Perhaps BLAMEWORTHINESS was too ambitious. Perhaps we aren't *always* in a position to know whether internalist conditions obtain but, still, there's a special kind of epistemic access we have to them that explains why they are especially well suited to determine blameworthiness. Let me briefly mention some ways of developing this strategy that don't appeal to luminosity, but that I think are ultimately unsuccessful.

First, one might claim that although we're not *always* in a position to know whether, for example, we're undergoing a certain experience, we're *usually* in a position to know what our experiences are like. The problem with this response is that it won't uniquely carve out internalist conditions. There are plenty of externalist conditions that we're *usually* in a position to know about: like, for example, the existence of medium sized objects in front of us in good lighting.

Second, one might claim that what's special about our access to internalist conditions is not that we're especially *reliable*, but rather that we have a special *way* of coming to know about internalist conditions: reflection, or introspection. The problem with this proposal is that, as is, it is not very explanatory. I have special ways of coming to know all sorts of things: for example, I have a special faculty for face recognition. But plausibly there isn't any interesting role in epistemology for norms whose antecedents describe only facts that I can detect with my face recognition capacity. So if it is the *way* of coming to know about internalist conditions that makes them especially well suited for, say, determining blameworthiness, much more would need to be said about why it is only conditions that one comes to know about through reflection or introspection that are relevant

⁴ For a philosophical argument against LUMINOSITY see Williamson (2000). For a response, see Berker (2008). For a response to Berker, see Srinivasan (2013). Schwitzgebel (2008) provides empirical arguments against claims like LUMINOSITY. I will not take a stand on LUMINOSITY here. My point is only to show that a version of internalism can be defended without it.

for blameworthiness (or whatever other roles the internalist norms are meant to play).

In sum, the motivation for hardcore externalism that is my focus is the following: The internalist thinks that there is some important role in epistemology that internalist norms are especially well suited to play. The externalist thinks that some explanation must be given for why externalist norms aren't equally well suited for the relevant role. Providing such an explanation naturally proceeds in two steps: first, a well-motivated constraint on the sorts of norms that are especially well suited to fulfill the role, and second, a reason for thinking that only internalist norms satisfy the constraint. The problem is that if LUMINOSITY is rejected, certain tempting constraints will rule out *all* non-trivial epistemic norms, whereas less demanding constraints will fail to exclude externalist norms. Thus, the internalist who doesn't want to rely on LUMINOSITY faces a challenge. This is the challenge I will address in the remainder of the paper.

3. What to Believe?

I suspect that, for many of us, our interest in normative epistemology began as an interest in *figuring out what to believe*. We wonder whether to trust the deliverances of perception, how to make inferences about the future from our knowledge about the past, how to respond to peer disagreement, what credences to assign if we find out that we'll be awoken on Monday if a coin lands Heads and we'll be awoken on Monday and Tuesday if it lands tails...and so on. What I will call the "what-to-believe" project is just the project of figuring out what to believe as a function of the circumstances that we find ourselves in. I will be assuming that we are aiming, in engaging in such a project, in settling on a way of forming beliefs that will aid us in the pursuit of truth.⁵ I will argue in what follows that internalist conditions play an important role in this project.

Karl Schafer (2014) has applied the term "doxastic planning" to describe the activity we engage in when we deliberate about what to believe. The reason for

⁵ Worries raised by Berker (2013) may lead some to doubt this claim. However, the idea that some connection must exist between epistemic rationality and the pursuit of truth is widespread. Addressing Berker's concerns is beyond the scope of this paper so I set them aside here.

thinking of this activity as a kind of *planning* is, I think, that the activity is aimed at settling on a course of belief formation, just as planning what to *do* is aimed at settling on a course of action.⁶ I might, for example, after thinking through the literature on peer disagreement, settle on suspending judgment about what's wrong with my car if I learn that my friend, an amateur mechanic herself, disagrees with me. Following Schafer, I too will describe the project of deliberating about and then settling on what to believe as the project of *doxastic planning*.

There is a sense, however, in which the planning terminology may be misleading. For given the purported involuntary nature of belief formation, one might think that there is no sense to be made of *planning* what to believe. So insofar as planning requires that the planned activity be under a kind of voluntary control that beliefs are not subject to, we shouldn't describe what we do when we think about what to believe as plan-making. But whether belief formation is voluntary or not, the following is clear: we can have an impact on what beliefs we form by considering how to respond to a body of evidence and then settling on one method of belief formation rather than another. Deliberating about what to believe, and whatever it is that happens when a conclusion of this sort of deliberation is reached, is all I mean to be talking about when I use the expression "doxastic planning." Formally, a doxastic plan will be represented as a set of conditionals of the form: *in circumstances C adopt doxastic state d*. (In other words, a plan is represented as a set of norms). An internalist plan will be a plan whose antecedents all describe internalist conditions, and I will call plans that are not internalist "externalist plans."

Schafer proposes that our attributions of *rationality* express our doxastic plans, and, by appealing to some constraints on planning, he argues that our attributions of rationality to an agent must depend exclusively on the set of internalist conditions that are true of that agent. I will argue in the next section that Schafer's position requires LUMINOSITY. There is, however, an alternative way to see the importance of internalist conditions for doxastic planning that does *not* appeal to LUMINOSITY. It is to this that I turn to now.

⁶ For further discussion of doxastic planning Schoenfield (forthcoming), Steel (ms.) and Greco and Hedden (ms.).

4. The Role of Internalist Plans in Doxastic Planning

In this section I will defend the claim that internalist plans are important because these are the plans that we should expect to *conform to* as a result of making whatever plans we make. I will begin with an example that illustrates the main idea:

PAUL REVERE: Paul Revere and his friends are considering how to disseminate information about the direction from which the British will be coming. Revere proposes the following:

The Revere Plan: To light one lantern if the British come by land, and two lanterns if the British come by sea.

Everyone agrees that this is a good plan. At a certain point, someone brings up the possibility that the British will use decoys to make it *appear* that they're coming by land, when in fact they're coming by sea, or make it *appear* that they're coming by sea, when in fact they're coming by land. The group starts taking this possibility seriously and, as a result, they begin to question the wisdom of the Revere Plan. But John Hancock is confused. He says:

“Why are we worried about whether there will be decoys? It's true that if there are decoys, it may *appear* to Revere that the British are coming by land, when in fact they're coming by sea. But Revere's plan doesn't involve lighting one lantern when it *appears to him* that the British are coming by land, and two lanterns when it *appears to him* that they're coming by sea. The plan is to light one lantern if the British *in fact* are coming by land and two lanterns if they *in fact* are coming by sea. So I don't see why the possibility of decoys is leading us to reconsider the original plan.”

It's clear how the group should respond: “We agree with you about what the plan recommends. However, we should expect that the result of *making* the Revere plan is that Revere will light one lantern if the British *appear* to him to be coming by land, and two lanterns if the British *appear* to him to be coming by sea (at least assuming that Revere is going to find out how the British are coming visually). Now, the reason we're making this plan in the first place is that it's important to us that accurate information be disseminated. So if we think that the result of making the Revere plan is that Revere will light one lantern if it appears that they're coming by

land and two if it appears they're coming by sea, we must assign (relatively) high expected value to Revere *conforming* to the following plan:

The Revere Plan:* To light one lantern if it *appears* that the British are coming by land, and two lanterns if it *appears* that the British are coming by sea.

If the likelihood of decoys is sufficiently high, conforming to the Revere* plan has low expected value. And so, if the likelihood of decoys is sufficiently high, there is a problem with *making* the original Revere plan."

There are two important features of this case worth noting: First, note that at no point did it matter whether Revere had introspective access to how things appeared to him. The reason that we should be interested in the Revere* plan is simply because we expect Revere to conform to Revere* as a result of making the Revere Plan. Introspection is irrelevant because we don't expect Revere to *first* form beliefs about how things appear to him, and *then* determine how many lanterns to light on the basis of introspection. We expect Revere to simply act on the basis of how things appear.

Second, nothing about the case suggests that there is any problem with Revere making the original plan when decoys are sufficiently unlikely, even though the plan is an externalist plan. So, I will not be arguing that there is anything wrong with making externalist plans. Rather, I will argue that our judgments about internalist plans constrain what plans are acceptable to make, be they internalist or externalist. This is in contrast with Schafer who aims to motivate internalism by arguing that there is something defective about *making* externalist plans.

The problem Schafer has with externalist plans is that these plans can require one to distinguish between indistinguishable circumstances. But Schafer's argument faces a dilemma. On a strong understanding of the constraint, it is implausible, and quickly leads to a LUMINOSITY commitment. On a weaker understanding, the constraint is plausible but it fails to rule out externalist plans.

On the strong understanding of the constraint one must *always* be able to distinguish between the circumstances that the plan distinguishes between. But note that the strong understanding rules out the following sorts of ordinary plans:

to go the store if there is no milk in the fridge, but stay home if there is milk in the fridge; to water the garden if it doesn't rain, but to not water the garden if it does rain. The strong version of the constraint rules out such plans because there are certainly *some* circumstances in which there is no milk in the fridge that I won't be able to distinguish from *some* circumstances in which there is milk in the fridge (for example, circumstances involving milk holograms). So the first problem is that the strong understanding of the constraint rules out even these ordinary plans. Perhaps one thinks that's not actually a problem. Really, one might claim, we should only plan to go to the store if it *appears* that there is milk in the fridge and not go to the store if *it doesn't appear* that there is milk in the fridge. But this proposal only helps if we assume LUMINOSITY. For on the strong understanding, one must *always* be able to distinguish between the circumstances that the plan distinguishes between. So the internalist plan above only satisfies the constraint if we are *always* in a position to know whether it appears that there is milk in the fridge. In sum, if Schafer's requirement only permits us to make plans when we can *always* distinguish between the circumstances that the plan distinguishes between, the constraint is implausible. It rules out a great number of ordinary and intuitively acceptable plans. If we reject the acceptability of such plans and retreat to internalist plans like the one above, this understanding of Schafer's constraint will require LUMINOSITY.

A more plausible understanding of Schafer's constraint would require only that we must *sometimes*, or *usually* be able to distinguish between the circumstances that our plans distinguishes between. But on this understanding there is no reason to rule out externalist plans. There are plenty of externalist conditions such that we can *sometimes* or *usually* determine whether they obtain (whether there is milk in the fridge, whether it is raining, and so on).⁷

⁷ Schafer claims that his defense of internalism is not committed to LUMINOSITY. But in arguing for this, he restricts his attention to what we might call *incomplete* plans, like the plan to water the garden if it doesn't rain, *while making no plans whatsoever for what to do if it does rain*. But as illustrated above we make plans to ϕ if C obtains and to ψ if $\sim C$ obtains all the time, so we certainly wouldn't want to add an *additional* constraint which would forbid us from making any plans of this sort.

For these reasons Schafer's argument does not successfully establish, without appealing to luminosity, that there is something defective about externalist plans. So rather than argue that internalist plans are the only doxastic plans that are appropriate to make, I will argue that internalist plans are the plans that we should expect to conform to as a result of making whatever plans we make. Why does this matter? After arguing for the claim, I will discuss its implications for the debate between internalists and externalists. The argument consists of two main premises, which I'll call "ENDORSEMENT" and "IMPLEMENTATION."

4.1. ENDORSEMENT

While some people might deliberate about what to believe for the sheer joy of it, I am assuming that epistemologists engaged in the what-to-believe project think of the project as aimed at a particular end: the pursuit of truth. Practical planning is also frequently aimed at bringing about some sort of desirable outcome. I don't usually plan my vacation for the fun of planning, but because I think that planning will bring about a better vacation than I would have if I didn't plan. With this in mind, I propose the following:

ENDORSEMENT: When the aim of planning is to bring about a valuable state of affairs, and you expect that the result of making a plan P is that you conform to a plan P', then, you should not make plan P unless you assign a (relatively) high expected value to conforming to P'.⁸

Why is ENDORSEMENT true? Simply because if the point of planning is to bring about a valuable outcome, and one believes that the result of making plan P is that one will

⁸ For uninteresting reasons, it will be simplest for what follows if we assume that the plans in question are *complete* – that is, they issue recommendations for all of the circumstances under consideration during planning. The argument I provide can be modified so as to apply to incomplete plans as well but it will be more cumbersome. Since the externalist plans that I'll be especially interested in are complete plans (believe p *if and only if* the belief that p would constitute knowledge, believe p *if and only if* such a belief is the product of a reliable process) this is a harmless restriction.

conform to P', it would be irrational to continue to endorse P while expecting that conforming to P' will lead to bad results.⁹

4.2. IMPLEMENTATION

IMPLEMENTATION: For any plan P that you make, there is some internalist plan P', such that you should expect that, as a result of making P, you will conform to P'.

Two clarificatory notes about IMPLEMENTATION: First, you needn't be *certain* that you'll conform to P' as a result of making plan P. My claim is just that you should *expect* that you will conform to P' as a result of making plan P. Second, the expectation here can be understood in a purely dispositional sense. I needn't consciously entertain any internalist plans when I'm making my plans. Revere and his friends didn't have to consciously consider the Revere* Plan while they were making the Revere Plan. However, the oddness of Hancock's comment, and the fact that the plan was given up immediately when decoys were made sufficiently likely suggests that, even before considering the decoys, the revolutionaries expected Revere to light one lantern if it appears to Revere that the British are coming by land and two if it appears that they are coming by sea.

Why accept IMPLEMENTATION? The case of Revere provides some intuitive support, but for a defense of the principle more generally, I will rely on Wedgwood's (2002) argument for the claim that our folk-psychological explanations of rule following always bottom out in internalist conditions. Wedgwood uses the following example: suppose someone adopts the rule to add salt to the water if the water is boiling. He says:

⁹ If there is a further plan that I'm aware of, P'', and I expect to execute P' by executing P'' then I am committed to assigning high expected value to that plan as well. But this will not go on indefinitely, since at a certain point there will simply be no further plan that I expect to conform to as a result of making the plan I make. See my discussion of Wedgwood in the next subsection.

If one follows this rule, then one's adding the salt is explained by, or is a response to, the fact that the water is boiling. However, it may be that the process whereby one's addition of the salt is explained by, or responds to, the fact that the water is boiling can itself be analysed, even at the folk-psychological level of explanation into a series of sub-processes. For example, perhaps the proximate explanation of one's attempt to add the salt is not the fact that the water is boiling, but rather one's belief that the water is boiling. Similarly, perhaps the proximate explanation of one's belief that the water is boiling is not the fact that the water is boiling, but one's having an experience that represents the water as boiling. In forming this belief in response to having this experience, one is following a rule. Specifically, one is following a rule that permits one to form the belief that the water is boiling when one has an experience that represents the water is boiling (and no special reason to distrust one's experience in the circumstances) (356).

Wedgwood goes on to argue that a fully articulated folk psychological explanation of *how* an agent comes to follow a rule (or implement a plan) with an *externalist* antecedent will appeal to the fact that the agent followed a rule with an *internalist* antecedent.¹⁰ This means that we can construct an internalist version of

¹⁰ Wedgwood uses this fact about our folk psychological explanations to motivate internalism about *rationality* and he too aspires to motivate the view without appeal to luminosity. I am not, however, convinced that he succeeds. This is because his strategy relies on the thought that the rules of rationality must be rules that can be followed directly. It is crucial to the argument that we understand the rules that can be followed directly as the rules that have as their antecedents conditions that our *folk-psychological* explanations bottom out in. But, like Srinivasan (note 12, this volume), I wonder what the motivation is for the claim that the rules of rationality are the rules that can be followed directly *in this sense*. The emphasis on our folk-psychological theory is especially worrying if, as Srinivasan suspects, "folk explanations bottom out at certain non-factive mental states because it is tacitly presumed that these states are transparent" (24). I share Srinivasan's suspicion, and I think her point can be illustrated by the following thought experiment:

Imagine that a society of Martians is discovered, and that these Martians have tacitly assumed that they would never make mistakes about the presence of medium sized objects in front of them: they take such states to be luminous to them. We know that they are mistaken. Martians can be envatted and deceived about the objects in front of them just like humans. Now note that it wouldn't be surprising if the Martian folk-psychological explanations bottomed out at the presence of medium sized objects, rather than internalist conditions. For the Martians would never feel the need to appeal to the fact that it *appeared to someone that there is a tree in front of them*, even though there isn't a tree, to explain an agent's beliefs or actions. (Indeed, if visual illusions of any sort *never*

any externalist plan by replacing the externalist antecedent with the conditions that would be appealed to in a fully articulated folk-psychological explanation of the action. Let me illustrate with an example: Suppose that Sally plans to add salt when the water is boiling, and let's stipulate that a fully articulated folk psychological of an attempted salt addition (as a result of adopting this plan) will appeal to the fact that it *appears* to Sally that the water is boiling. Then, we should expect that, as a result of adopting her plan, Sally will attempt to add salt when it *appears* that the water is boiling.¹¹

Note that the reason we should expect to conform to internalist plans as a result of making externalist plans is not that we have some special epistemic access to internalist conditions. Rather, it is because internalist conditions are more *causally proximate* to the action or belief in question than externalist conditions are. If Billy's throwing the ball causes the window to break, which causes shards of glass to fall into the room, then whether the window breaks is a better predictor of whether shards of glass will fall into the room than whether Billy throws the ball.

occurred, it wouldn't be surprising if the Martians didn't even develop a term for it *visually seeming to one as if p*).

Now, take some rule that Wedgwood regards as a rule of rationality: ϕ if and only if C obtains. Suppose it turns out that C might obtain but some agents intending to follow this rule wouldn't ϕ because they wouldn't enter an intermediate psychological state, C' , that causally mediates between C and ϕ -ing, which we might describe as *representing that C obtains*. Our own folk-psychology never developed a term for *the representing that C obtains* because, perhaps, we think (correctly or incorrectly) that agents who adopt the rule to ϕ if and only if C obtains, are always in a position to do just that. If such a C' exists, whatever motivates the thought that the rules of rationality are rules that agents can follow *directly* would, I think, motivate the thought that a rational agent will ϕ if and only if C' obtains and not the thought that a rational agent will ϕ if and only if C obtains. This is so despite the fact that our folk psychological explanations, for some contingent reasons, never developed a term referring to C' . To see this, note that any internalist worth her salt should want to claim that an envatted Martian is rational, even though the *Martian's* folk psychological explanations bottom out in non-internalist conditions. If we think that the rules of rationality must be rules that can be followed directly, we will think that the envatted Martians are rational because *in fact* the only rules they can follow directly are rules with internalist antecedents. We don't care what rules *they* think can be followed directly. So if *our* folk psychology turns out to be wrong about which rules we can follow directly, just like the Martians' folk psychological theory is wrong, why should we think that all agents are rationally required to conform to rules that have as their antecedents conditions that our mistaken folk psychology bottoms out in? My proposal will be able to better explain the relevance of folk-psychology. But my internalist ambitions are also more modest than Wedgwood's.

¹¹ Wedgwood doesn't deny Williamson's claim that knowledge attributions may play a role in explaining certain kinds of action, but he argues that factive mental states like knowledge cannot be part of the fully articulated folk-psychological explanation of non-factive mental states like beliefs. See his paper for the details of the argument.

Similarly, if water boiling causes appearances as of water boiling, which causes beliefs that water is boiling, then whether it appears as if water is boiling is a better predictor of whether a belief will be formed than whether the water is actually boiling. Thus, if Wedgwood is right that folk-psychological explanations of rule following bottom out in internalist conditions, then for any plan one makes, there will be some internalist plan we should expect an agent to conform to, as a result of making the plan. Is it problematic that the argument for IMPLEMENTATION relies on our empirically informed folk-psychological theories? I will address this question in section 6.

4.3. E-LINK

It follows from ENDORSEMENT and IMPLEMENTATION that:

E-LINK: When the aim of planning is to bring about a valuable state of affairs, then, for any plan P that you make, there is some internalist plan P' such that, you should not make plan P unless you assign a (relatively) high expected value to conforming to P'.

E-LINK tells us that the expected value that we assign to conforming to internalist plans substantially constrains what plans are acceptable plans for us to make when the aim is bring about a valuable state of affairs. In the what-to-believe epistemology project that I've been describing, the aim *is* to bring about a valuable state of affairs: one in which we have true beliefs and avoid false beliefs.¹² It follows that, from the perspective of what-to-believe epistemology, we should only adopt a doxastic plan if we assign high expected value to conforming to the internalist version of that plan. This is why internalist plans play an important role in the what-to-believe project.

5. Upshots for the Internalism/Externalism Debate

¹² Or, in the case of credences, the aim is to achieve a high accuracy score as measured by some scoring rule.

For epistemologists engaged in the what-to-believe project, E-LINK has some important consequences.

(a) Externalist Solutions to Skepticism

The skeptical problem arises from the uncomfortable realization that things might be exactly the same “from the inside” while the facts outside vary wildly. The externalist’s response is to suggest that it is a mistake to assume that it is only how things are “from the inside” that is relevant to what we should believe. In making this move it seems that the externalist gets the prize (beliefs!) without having to do the hard internalist work of justifying the likelihood of certain hypotheses that are undetermined by how things are from the inside. It also seems that the externalist is offering a genuine alternative to the internalist’s proposal. But if I take the externalist’s recommendations as my plan for what to believe, I should expect the result to be that I conform to some internalist version of that plan. If I plan to believe *p* tomorrow if the belief would be the result of a reliable process, I should expect, as a result of making this plan, that I will believe *p* tomorrow if I *believe* or *judge* that the relevant process is reliable. So, as far as skeptical worries go, it doesn’t look like externalist plans have much of an advantage over their internalist cousins. Whatever worries one might have about forming beliefs based on how things appear from the inside are worries that will be shared by the externalist planner. This is because she should expect to do *exactly that* as a result of making her externalist plan.

(b) The Antecedent Unlikelihood of Skeptical Scenarios

The Revere case illustrates that, frequently, when we make plans, we are committed to thinking, *prior to the circumstances in question arising*, that it is unlikely that we’ll find ourselves in a skeptical scenario. Revere should only make his original plan if, *when making the plan*, he thinks it is unlikely that his experiences will be misleading.

A similar phenomenon arises when we make plans for what to believe. Suppose I’m considering what to believe about the color of the wall in a room that I

will be walking into. I know that the only evidence I will receive is that which will result from my looking at the wall. Many internalists think that my relevant evidence in such a case will be something like *the wall's appearing red to me* (and, perhaps, the absence of defeaters). On many of these views,¹³ for it to be rational for me to believe that the wall is red upon entering the room, it must be rational, *prior* to entering the room, to think it's likely that the wall is red *conditional on it appearing red* (and the absence of defeaters). But some externalists (I'm thinking particularly of "Knowledge-Firsters") will claim that if, upon entering the room, I come to *know* that the wall is red, then *that the wall is red* is part of my evidence.¹⁴ So for my belief to be justified upon seeing the wall, it's not necessary that it is rational, prior to seeing the wall, to think that the wall is likely to be red if it *appears red* (and there are no defeaters). After all, my evidence, if I come to know that the wall is red, won't be exhausted by the claim that the wall *appears* red. It will also contain the proposition that it *is* red!¹⁵

Now, consider two doxastic planners: The internalist says: "I'm going to believe that the wall is red if and only if it appears red and there are no defeaters." The externalist says: "I'm going to believe that the wall is red if and only if such a belief would constitute knowledge." Nothing I have said suggests that either of these plans is better than the other. However, if the externalist planner expects to form beliefs about the color of the wall upon entering the room, then she *too* is committed to thinking, prior to entering the room, that *it's likely that the wall will be red conditional on it appearing red (and there being no defeaters)*. This is because she should expect that, as a result of making her externalist plan, she will end up believing that the wall is red *in exactly those circumstances in which the internalist plan recommends such a belief*. (I imagine that, despite their disagreements in

¹³ Jim Pryor's (2000) "dogmatism" is a notable exception.

¹⁴ Some varieties of dogmatism, such as those described in Miller (ms.) and Moretti (2015), endorse a related view: that if I have an experience as of a red wall (and no defeaters), then *that the wall is red* is part of my evidence, whether it is red or not (they therefore must allow that one's evidence can contain false propositions). The argument below shows that, as far as the what-to-believe project goes, such views are *also* committed to assigning a high antecedent likelihood to p being the case conditional on it appearing as if p.

¹⁵ Maria Lasonen Aarnio (2010) goes even further, claiming that not only isn't a *prior* antecedent likelihood of reliability necessary, the knowledge can be retained even in the presence of defeaters.

epistemology, Timothy Williamson and Richard Feldman largely agree about the color of walls they have both encountered!). The bottom line is that, whatever plans externalists make, they too will be duped in skeptical scenarios. So if they are to continue to endorse their plans, they must, in the planning phase, regard such scenarios as antecedently unlikely.

(c) Epistemic Praise and Blame

Here are the central claims I've argued for thus far: First, one way of thinking about the project that (some) epistemologists are engaged in is as *deliberating about what to believe*. And second, the acceptability of a plan for what-to-believe is constrained by the desirability of conforming to an internalist version of that plan. Plan P is a good plan for what to believe only if we can antecedently expect good epistemic outcomes if we conform to the internalist version of P. This suffices to show that internalist plans play an important role in the what-to-believe project.

What I haven't done so far is say of any particular *status* that whether it obtains supervenes on a set of internalist conditions. That is, I haven't argued that the status of an agent as justified, rational, or epistemically blameworthy is determined by a set of internalist conditions. I will, below, suggest a way in which the arguments I've given for the role that internalist conditions play in *planning* might be deployed in an argument for the claim whether an agent is *epistemically blameworthy* should be determined by a set of internalist conditions.

However, before providing the argument I would like to emphasize that, for a what-to-believe epistemologist, the question of whether some epistemic *status* is determined by internalist or externalist conditions is not the question surrounding the internalism/externalism debate that is of primary interest. To see this, note that if you're interested in making a plan for your vacation, you won't usually spend a lot of time thinking about whether various ways in which you might implement your plan should be praised or blamed, should count as conforming to the plan or should not count as conforming to the plans, and so on. This is because you are primarily interested in choosing plans that you expect will bring about *good vacations*; not plans that you expect will bring about some status such as "having conformed with

one's plan" or "having implemented one's plan in a way deserving of praise." Similarly, the epistemologists I have in mind that are interested in deliberating about what to believe are interested in choosing plans that they expect will bring about *true beliefs*, not plans that will bring it about that one is to be praised or to be given some sort of epistemic thumbs-up.

What *is* of interest to what-to-believe epistemologists concerning the internalism/externalism debate are the following two questions:

(Q1) Should all of our plans for what-to-believe be internalist plans? Or is it okay to make externalist plans?

Schafer has claimed, in response to Q1, that our plans for what-to-believe should all be internalist plans. I have argued that this restriction is inconsistent with ordinary practice and that Schafer's argument for this restriction relies on luminosity. So I don't think that internalism is vindicated in the sense that *all epistemic plans should be internalist plans*.

Here is the second question surrounding the internalism/externalism debate that will be of interest to a what-to-believe epistemologist:

(Q2) Is there any important role that internalist plans should play in deliberation about what-to-believe?

The answer to this question is *yes*. What we expect to result from the internalist versions of our plans constrains what plans are acceptable to make.

Despite the fact that I think a what-to-believe epistemologist should be primarily interested in Q1 and Q2, and not whether certain epistemic behaviors are deserving of praise or blame, it may be possible to develop the arguments I've given about planning into an argument for internalism about blameworthiness and praiseworthiness. Below, I will roughly describe one way of doing so that takes as its starting point a proposal from Sinan Dogramaci (2012) about the function of

epistemic praise and blame.¹⁶ I do not wish to take a stand on the proposal itself. My aim is just to illustrate a possible way of extending the arguments I've given about planning to the domain of praiseworthiness and blameworthiness.

Dogramaci proposes that the *function* of epistemic praise and blame is to modify the epistemic behavior of others. Why do we want to modify the epistemic behavior of others? Because, Dogramaci argues, we are not epistemic islands: much of our epistemic life involves relying on one another's testimony. Thus, we have an interest in promoting epistemic coordination. I want you to form beliefs the way I do so that I can rely on your testimony in the future. I think that *if* Dogramaci is right that the function of epistemic praise and blame is to modify the epistemic behavior of others, then the claims I've made so far support internalism about epistemic blameworthiness and praiseworthiness.

Here's the thought: in paradigmatic cases in which we blame an agent for ϕ -ing in circumstances C, and the blame impacts the person's behavior in the desired way, the agent, upon being blamed, makes some modification that she expects will lead to different behavior if circumstances C arise again. For example, if I'm blamed for arriving late to a faculty meeting, what one would expect me to do, if all goes well, is modify my behavior in ways that I expect will make it more likely that I *don't* arrive late to faculty meetings. (Perhaps, if such a meeting is scheduled, I'll set my alarm clock for an earlier time, eat a shorter breakfast, or the like). It would be odd for me to react to the blame by modifying my behavior in some way that I would expect to lead to the very same results (arriving late) in the very same circumstances (scheduled faculty meetings). This may not be an exception-less rule – it is sufficient that, in paradigmatic cases, blame, if successful, leads the agent to make a modification that she can expect will lead to a different outcome if similar circumstances arise.

Let's now imagine a case in which P is the plan that we want agents in our community to adopt. So as not to beg the question against the externalist, and to make things as difficult as possible for the internalist, let's suppose that P is an

¹⁶ For further developments see Dogramaci (forthcoming) and Dogramaci (ms.) For a similar proposal applied to the ethical sphere see Miranda Fricker (forthcoming).

externalist plan. I have argued that what we should expect will happen if an agent adopts P is that she'll conform to some internalist plan. The same arguments support the claim that what we should expect will happen if an agent adopts P *and tries her utmost to follow P* is that she'll conform to some internalist plan – call it P'.

Now suppose that an agent violates P (the externalist plan we want our agents to adopt) but conforms to P' (the internalist plan that we'd expect agents who have adopted P, and are trying their utmost to follow P, to conform to). Should we criticize such an agent? No! For recall that the paradigmatic way in which blame impacts an agent's behavior is that the agent, when criticized, will make a modification that she can expect will lead to a different outcome in similar circumstances. This means that we should only criticize an agent for a behavior if we *want* the agent to make a modification that she can expect will lead to a different behavior in similar circumstances. But the only thing that an agent who violated P but conformed to P' can do, which she can reasonably expect will lead to a different outcome in similar circumstances, is *adopt a plan other than P* (or adopt no plan at all). For adopting P, and even trying her utmost to follow P, will lead to exactly the behavior that she was just criticized for in exactly the circumstances that just arose. Thus, we should only blame the agent who conforms to P' but violates P if we don't want her to adopt P. But, by stipulation, we *do* want her to adopt P! Thus, we shouldn't blame an agent who conforms to P' but violates P.

Analogous reasoning shows that we shouldn't praise an agent who conforms to P but violates P'. This means that, insofar as P and P' come apart, if we want agents to adopt P, whether we should praise or blame an agent will depend on whether the agent's behavior is in conformity with P'. It follows that even if the plan we want agents to *adopt* is an externalist plan, whether we praise or blame an agent should be determined by whether the agent conforms to *the internalist version* of that plan. It follows that, on Dogramaci's picture, whether an agent should be epistemically praised or blamed will be determined by a set of internalist conditions.

In sum, there are three important lessons to be drawn from E-LINK that are relevant to debates between internalists and externalists. The first is that, for a what-to-believe epistemologist, externalist theories fare no better than internalist

ones as far as alleviating skeptical concerns go. Since we expect the result of making an externalist plan to involve conforming to an internalist plan, whatever underdetermination worries one might have had about internalist plans are worries that the externalist planner will share. Second, if one expects to form beliefs on the basis of perception, both internalists *and* externalists making plans about what to believe are committed *antecedently* to assigning a high probability to p conditional only on it appearing to one as if p. Finally, if the function of assigning epistemic praise or blame is to promote coordination, then whether an agent is epistemically praiseworthy or blameworthy will be determined by a set of internalist conditions. This is so even if we want the plan that agents in our community adopt to be an externalist plan.

6. Conclusion: Is it Internalist Enough?

I've argued that internalist plans play an important role in our deliberation about what to believe. My argument for this claim relies on our expectation that, as a result of making any sort of plans, we'll conform to internalist plans. But this expectation of ours is about an empirical matter: the way we expect our psychology to operate in the future. So one might wonder: if we were conducting the what-to-believe project *purely a priori* would we have any special reason to be interested in internalist plans? This may seem like an especially pressing question since one might think that the internalist *project*, at least as originally conceived by Descartes, involved exactly that: figuring out what to believe *without making any assumptions about the past, the future, or the external world*.

If the what-to-believe project is meant to be carried out with absolutely no assumptions about the past, future, or the external world, then, I fear, the justification I have offered for the importance of internalist plans will not be available. However, I want to suggest that it would simply be a mistake to think that *no* beliefs of this sort should be relied on in such a project – even if one is engaged in a Cartesian sort of what-to-believe project. This is because a prerequisite for sensibly engaging in *any* sort of planning, or settling, is that one have *some* expectations about the future which make it reasonable to predict that the mental

act one is engaged in (the planning or settling) will yield certain sorts of outcomes.¹⁷ This means that, at minimum, some beliefs (which need not be consciously entertained) about *psychology* are required for planning. For even if the plan only involves plans to engage in mental acts when certain internalist conditions are satisfied (like to *believe* *p* in internalist condition X, or to *try* to raise one's arms in internalist condition Y) the subject needs to think that whatever mental act she is engaged in when settling on a plan will likely bring about some sort of mental activity in the vicinity of what the plan recommends. Without any such expectation it's unclear what about the agent's psychology would make the act in question a *plan* rather than merely, say, a hope or a desire.

So suppose that, like Descartes, we want to rid ourselves of as many beliefs as possible, and engage in the what-to-believe project from a very minimal basis.¹⁸ Since, I have argued, we must start with *some* expectations about psychology to sensibly engage in this project, it is natural to start with our *actual* psychological beliefs concerning what will result from settling on one way of forming beliefs rather than another. According to our folk-psychology, what we should expect from an agent who adopts an externalist plan is that she will conform to an internalist plan. If our psychological theory changes in the future, and internalist plans are no longer what we expect to conform to as a result of making whatever plans we make, then, it may turn out, internalist plans will no longer play an important role in deliberating about what-to-believe. But given the way that we actually conceive of our future actions and future belief-formations, what *we* expect will result from conforming to internalist plans should constrain what externalist plans we make.¹⁹

¹⁷ Although it is controversial exactly *what* empirical beliefs about the future one needs in order to engage in planning, it is commonly accepted that *some* such beliefs are required. See for example Harman (1976) and Bratman (1987).

¹⁸ Note that I'm not suggesting here that this *is* the most fruitful perspective from which to engage in the what-to-believe project. (Though I do have some sympathy to this idea. See my (forthcoming)). I am just suggesting that we imagine engaging in the project from this perspective since this is the sort of project that the internalist pressing this worry is interested in.

¹⁹ Note how my account deals with the worry that I raised for Wedgewood's argument in note 10. If Martians tried to engage in a Cartesian-like what-to-believe project, but they hadn't developed concepts that describe internalist conditions, they would end up engaging in the what-to-believe project using their *own* psychological theory. Thus, they would predict that, as a result of making their plans, they will conform to some set of plans that have as their antecedents conditions about the

Crucially, for my purposes, the importance of internalist conditions in deliberating about what to believe ultimately turns on the important *causal* role that we expect them to play: they need not be privileged epistemically.

References

- Alston, W. (1993). "Epistemic Desiderata." *Philosophy and Phenomenological Research* 53 (3): 527-551
- Alston, W. (2005). *Beyond Justification: Dimensions of Epistemic Evaluation*. Cornell University Press.
- Berker, S. (2008). "Luminosity Regained." *Philosophers' Imprint* 8 (2): 1-22.
- Berker, S. (2013). "Epistemic Teleology and the Separateness of Propositions" *Philosophical Review* 122: 337-393.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. CSLI Publications.
- Descartes, R. (1853/1996). *Meditations on First Philosophy*. Tr. John Cottingham. Cambridge University Press.
- Dogramaci, S. (2012). "Reverse Engineering Epistemic Evaluations." *Philosophy and Phenomenological Research* 84(3): 513-30.
- Dogramaci, S. (forthcoming). "Communist Conventions in Deductive Reasoning." *Noûs*.
- Dogramaci, S. (ms.). "Forgive and Forget."
- Foley, R. (2001). *Intellectual Trust in Oneself and Others*. Cambridge University Press.
- Fricker, M. (forthcoming). "What's the Point of Blame?" *Noûs*
- Goldman, A. (1979). "What is Justified Belief?" In George Pappas (ed.), *Justification and Knowledge*. D. Reidel: 1-25.

medium sized objects in front of them. That's okay. The reason it turns out that our own psychological theory is important isn't that we are certain that it is correct, or that the rules we're rationally required to conform to depend on which psychological theory we happen to adopt, but rather because we need some beliefs about what we can expect as a result of making plans if we want to engage in planning to begin with, and we will, inevitably, use the beliefs about this matter that we find ourselves with.

- Goldman, A. (1988). "Strong and Weak Justification." *Philosophical Perspectives* 2:51-69.
- Goldman, A. (1999). "Internalism Exposed." *Journal of Philosophy* 96(6): 271-293.
- Greco, D. and Hedden, B. (ms.). "Uniqueness and Metaepistemology."
- Greco, J. (2005). "Justification is Not Internal." In Steup M. & Sosa E. (eds.), *Contemporary Debates in Epistemology*. Blackwell: 257-269.
- Harman, G. (1976). 'Practical Reasoning,' reprinted in *The Philosophy of Action*, A. Mele (ed.), Oxford University Press: 149-177.
- Jacobson, S. (1997). "Externalism and Action-Guiding Epistemic Norms." *Synthese* 110 (3): 343-355.
- Kornblith, H. (1988). "How Internal Can You Get?" *Synthese* 74(3): 313-327.
- Lasonen Aarnio, M. (2010). "Unreasonable Knowledge." *Philosophical Perspectives* 24: 1-21.
- Miller, B. (ms.). "How to be a Bayesian Dogmatist."
- Montmarquet, J. (1993). *Epistemic Virtue and Doxastic Responsibility*. Rowman & Littlefield Publishers.
- Moretti, L. (2015). "In Defense of Dogmatism." *Philosophical Studies* 172(1): 261-82.
- Plantinga, A. (1993). *Warrant and Proper Function*. Oxford University Press.
- Pryor, J. (2000). "The Skeptic and the Dogmatist." *Noûs* 34:4 517-549.
- Pryor, J. (2001). "Highlights of Recent Epistemology." *The British Journal for the Philosophy of Science* 52(1): 95-124.
- Schafer, K. (2014). "Doxastic Planning and Epistemic Internalism." *Synthese* 191 (12): 2571-2591.
- Schoenfield, M. (forthcoming). "Bridging Rationality and Accuracy." *The Journal of Philosophy*.
- Scwitzgebel, E. (2008). "The Unreliability of Naïve Introspection." *Philosophical Review* 117 (2): 245-273.
- Sosa, D. (forthcoming). "The Unimportance of Being Earnest." *Philosophy and Phenomenological Research*.

- Sosa, E. (2001). "Reliabilism and Intellectual Virtue." In H. Kornblith (ed.), *Epistemology: Internalism and Externalism*. Blackwell Publishers.
- Steel, R. (ms.). "Peer Disagreement, Anticipating Failure and Avoiding It."
- Srinivasan, A. (2013). "Are We Luminous?" *Philosophy and Phenomenological Research* 89(2): 294-319.
- Srinivasan, A. (this volume). "Normativity without Cartesian Privilege."
- Steup, M. (1999). "A Defense of Internalism." In L. Pojman (ed.), *The Theory of Knowledge: Classical and Contemporary Readings, 2nd edition*. Wadsworth Publishing.
- Wedgwood, R. (2002). "Internalism Explained." *Philosophy and Phenomenological Research* 65(2): 349-369.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press.