

# Radical epistemology, structural explanations, and epistemic weaponry

Richard Pettigrew\*

May 9, 2021

## Abstract

When is a belief justified? There are three families of arguments we typically use to support different accounts of justification: (i) arguments from our intuitive responses to vignettes that involve the concept; (ii) arguments from the theoretical role we would like the concept to play in epistemology; and (iii) arguments from the practical, moral, and political uses to which we wish to put the concept. I focus particularly on the third sort (iii), and specifically on arguments of this sort offered by Clayton Littlejohn (2012) and Amia Srinivasan (2018) in favour of externalism. I counter Srinivasan's argument in two ways: (a) first, I show that the internalist's concept of justification might figure just as easily in the sorts of structural explanation Srinivasan thinks our political goals require us to give; and (b) I argue that the internalist's concept is needed for a particular political task, namely, to help us build more effective defences against what I call epistemic weapons. I conclude that we should adopt an Alstonian pluralism about the concept of justification.

When is a belief justified? While it might well be true, as Dutant (2015) claims, that few philosophers ever actually advocated the analysis of knowledge as justified true belief, Gettier's (1963) brief refutation of it was undoubtedly the catalyst for an enormous effort by epistemologists to understand when a belief does count as knowledge, and of course to do that they had to understand when a belief counts as justified, since most of them agree that justification is required for knowledge. Since then, many putative answers to the latter question have been proposed, and one of the central points of disagreement that has emerged is between internalists and externalists.

---

\*I would like to thank the following people for helping me to think through these questions: Clayton Littlejohn, Amia Srinivasan, Dan Greco, Liam Kofi Bright, Catrin Campbell-Moore, James Ladyman, Zoe Johnson King, and Josh Habgood-Coote. I would also like to thank two anonymous reviewers.

The debate between these two camps is often set up with internalism as an extreme position that says that only internal states of a subject are relevant to whether their beliefs are justified, while externalism is just the negation of internalism, covering any view that takes any external state to be in any way relevant. But I think it's more illuminating to array the various positions across a landscape that has extreme internalism at one edge of the map, extreme externalism at the opposite edge, and all other positions spread out in between them. Where you lie in this landscape is determined by the extent to which the conditions that suffice for and are required for justification concern matters internal to the subject who has the belief and the extent to which they concern matters external to them. Perhaps the most extreme internalist position says that whether or not you are justified in believing something at a given time depends only on those of your mental states that occur at that time and that you actually access at that time. And the most extremely externalist position says that a belief is justified just in case it is true. In between, we have a large array of positions. We have what Conee & Feldman (2001) call *mentalism*, which allows mental states that are not only not actually accessed but also not accessible to help determine whether a belief is justified. And we have *process* and *indicator reliabilism*, both of which appeal to an internal state—the belief-forming process, or the grounds of the belief—but consider mainly external features of those internal states—the proportion of the beliefs that the process produces that are true, or the objective probability that the belief is true given that the subject has that ground for it (Goldman, 1979; Alston, 1988). And there are many more.

In this paper, I am interested in how we might choose between the positions in this landscape. The clearest inventory of arguments for and against different versions of internalism and externalism about justified belief is found in the introduction to Clayton Littlejohn's *Justification and the Truth-Connection* (Littlejohn, 2012, 1-61). The arguments he enumerates there can be divided into three categories based on the considerations they adduce in favour of the various accounts: in the first sort of argument, we appeal to the intuitive responses that competent users of the concept give when they are presented with particular cases; in the second, we appeal to theoretical considerations from within epistemology; and in the third, we appeal to the practical uses to which we would like to put the concept, including moral and political uses. In this paper, my interest lies in arguments from the third category.

Before we begin, it's worth noting that each of these three sorts of argument targets a slightly different conclusion. Arguments of the first sort are given in the service of a descriptive project. The aim is to map the borders of our existing concept of justification, assuming of course that there is a coherent concept to be mapped. Arguments of the second sort are also concerned with that existing concept, but they are open to modifying it,

perhaps quite substantially, in order to make it better fit into the theoretical role we have written for it in our epistemology. Arguments of the third sort are much less concerned with the existing concept, and much more concerned with whatever concept might serve our practical, moral, and political ends best. As a result, the three approaches could deliver three different accounts of justification without thereby being incompatible. It might be that the concept we actually use is not the one optimised for theoretical purposes, and that neither of these is the most useful practically, morally, and politically.

In this paper, I'm particularly interested in the third family of arguments. Arguments of this sort appeal to the practical use to which we wish to put the concept of justification, including its moral and political uses. Before we meet some of these arguments, it is worth considering how a concept might be put to political use. In fact, there are many ways. Concepts help us to group together similar phenomena. This might allow us to identify a regularity in observed phenomena and infer inductively that it will continue in unobserved phenomena; or it might allow us to recognise that all things of a certain sort should be treated morally or politically in the same way; and so on. So, for instance, a concept might help us politically if we value or disvalue for the same reasons everything that falls under it. Dotson's concept of an epistemic oppression might be an instance of this, as might Frye's concept of a double bind or Collins' concept of a controlling image (Dotson, 2012; Frye, 1983; Collins, 2000). Before we adopt the concept of epistemic oppression, we might notice in many disparate individual instances of it that something bad has happened—the boundaries of a debate we witness are circumscribed early on in a way that prevents certain groups from contributing; access to certain repositories of knowledge or certain modes of education is policed in a way that masquerades as objective, but in fact serves to exclude certain groups without good reason; and so on. But by grouping these instances together under a single concept, we can think about them systematically. And, if we agree that they are all bad, we can investigate how to structure our society in ways that reduce their prevalence. So, our political goals are furthered by adopting the concept.

Another way a concept might help us politically: it might help us to identify a regularity in what we have observed so far, and then allow us to infer inductively that this regularity will continue into the future. The concept of the Overton window might be an instance of this. We might have noticed in the past that more minor politicians from a ruling political party have started to espouse extreme positions on some issue shortly before their party announces a policy that is less extreme than these functionaries have espoused but more extreme than what the country they govern would previously have tolerated. We then adopt the concept of the Overton window and the next time we hear a governing party defending more extreme views than we expect, we infer that they are deliberately

shifting the Overton window to smooth the passage of a policy that they will shortly announce. Again, our political goals are furthered, because we can better understand and predict the behaviour of the politicians whose decisions affect us.

So these are some of the ways in which a concept might help us politically. Now, let's begin our discussion of the third family of arguments for different accounts of justification. The most well known from this family appeals to our use of epistemic concepts like justification when we assign blame to someone, either informally or in the context of the law. Jeremy picks up what looks exactly like his bottle of gin from exactly the spot where he'd put down his bottle only an hour earlier; he pours Isaak a glass, and Isaak drinks it; unbeknownst to both, Felix had earlier put Jeremy's gin back in the cupboard and absent-mindedly left his bottle of paint stripper on the worktop instead in a bottle identical to Jeremy's, and that is what Jeremy poured for Isaak. Isaak becomes severely ill as a result (Williams, 1981, 102). Is Jeremy to blame? To answer that, we might naturally ask whether Jeremy's belief that the bottle contains gin is justified. If it is, we might judge him innocent; if it is not, we might not be so lenient. As in the case of epistemic blame discussed in the previous paragraph, we might then reason that whether or not I can be blamed for acting on the basis of a belief I have can only depend on features of that belief that are accessible to me. I cannot be blamed if, by some bad luck, there is an external feature of the belief—its falsity, for instance—that was not accessible to me. Therefore, the argument concludes, internalism.

Clayton Littlejohn (2012) also offers an argument that belongs to this third category. And he too is interested in what the concept of justification must be like if it is to play the role we'd like it to play in our normative theorising about actions.

I think we should [...] try to understand what is involved in justification by trying to understand what is involved in properly relying on a belief for the purposes of practical deliberation. (Littlejohn, 2012, 199)

But he draws an externalist conclusion, not an internalist one. Indeed, most distinctively, he concludes that justification is a factive concept: if a belief is justified, it is true. And he reasons to that conclusion by arguing that a belief cannot justifiably be included in practical deliberation unless it is true, and it cannot be justified unless it can justifiably be included in practical deliberation. But what of Jeremy's belief that he is pouring a glass of gin for Isaak? Was that not justifiably included in his practical deliberations? For Littlejohn, it was not. It may be reasonably or blamelessly or excusably included therein, but not justifiably. And indeed Littlejohn is careful to make room throughout for a notion of blameless or reasonable or excusable belief, and blameless or reasonable or excusable actions based on such

states. But, like Austin (1956), he insists these are different from the notion of justified beliefs and the justified actions based on those states. Indeed, if I read him right, he even accepts that it's correct to say that Jeremy is *personally justified* in believing that there is gin in the glass; what it is wrong to say is that Jeremy's belief is *doxastically justified*. So there is some sense in which the belief is justified—it is the belief of a subject who is personally justified in holding it. But it is not itself doxastically justified (Littlejohn, 2012, 59).

Let's consider one of Littlejohn's central examples to see how the argument works. He presents two versions of his LOAN SHARK case (Littlejohn, 2012, Section 6.4.3). In each, a man approaches Harry. Harry believes that the man is his nemesis, Bobby, who is intent on harming him. The man looks exactly like Bobby. Harry pulls out a revolver and takes aim. In the first version of the case, Harry's belief is true, while in the second it false—it is Bobby's identical twin brother who is approaching Harry, and while he looks mean, he has no ill-intent towards Harry. In the first but not the second, Littlejohn thinks that Harry retains his right to non-interference from a third party. That is, if Audrey were to use force to prevent him from shooting the approaching man, he would have cause for complaint only in the first case. And this, we might think, is because he was going to commit a justified harm to the man in the first case, because it was an act of self-defence, but an unjustified harm in the second case, because, though Harry took it to be an act of self-defence, it wasn't. And in order to deliver this result, we must ensure that our concept of justification renders the individual's belief justified in the first case, but not in the second. Therefore, externalism.

So far, then, we've met two arguments from this third category. In both cases, they are concerned with the use to which we put the concept of justification for a belief when we are evaluating the normative status of an action that is based on it. They come to opposite conclusions. I won't try to adjudicate the dispute here, not least because I will conclude later that there is no need—both can be right. Instead, I turn to another argument from the third category, which is due to Amia Srinivasan (2018). Although it does belong to this third category, it begins with an argument from the first category. That is, Srinivasan presents three vignettes, asks us to agree in each with her intuitive evaluation of a particular belief held by its protagonist, and notes that only the externalist can vindicate these intuitive verdicts. Here are the vignettes:

**RACIST DINNER TABLE** Nour, a young British woman of Arab descent, is invited to dinner at the home of a white friend from university. The host, Nour's friend's father, is polite and welcoming to Nour. He is generous with the food and wine, and asks Nour a series of questions about herself. Everyone laughs

and talks amiably. As Nour comes away, however, she is unable to shake the conviction that her friend's father is racist against Arabs. But replaying the evening in her head she finds it impossible to recover just what actions on the host's part could be thought to be racist, or what would justify her belief in the host's racism. If pressed, Nour would say she "just knows" that her host is racist. In fact the host is racist – he thinks of Arabs as inherently fanatic, dangerous and backwards – and as a result did send off subtle cues that Nour subconsciously registered and processed. It is this subconscious sensitivity that led to her belief that her host is racist. (Srinivasan, 2018, 2)

CLASSIST COLLEGE Charles is a young man from a working-class background who has just become the newest fellow of an Oxford college. He is initially heartened by the Master's explicit commitment to equality and diversity. The Master assures him that, though the college is still dominated by wealthy fellows, Charles will be welcomed and made to feel included. Indeed, the Master tells Charles, he too is from a working-class background, and has experienced plenty of discrimination in his time. Charles is confident not only that the college will be a good community for him, but also that the Master is a person of excellent judgment on these matters. However, a few incidents soon disrupt Charles's rosy view of things. At high table, when Charles explains that he went to a state school, a fellow responds with 'but you're so well-spoken!'. At a visit to the pub, a number of young fellows sing the Eton boating song while Charles sits uncomfortably silent. Finally, Charles hears that the other fellows call him "Chavvy Charles's". Charles, who has a dependable sensitivity to classism, goes to the Master to report that he has experienced a number of classist incidents in college. Shocked, the Master asks him to explain what happened. But when Charles describes the incidents, the Master is visibly relieved. He assures Charles that none of these are genuinely classist incidents, but playful, innocuous interactions that are characteristic of the college's communal culture. He tells Charles that he is sure that Charles himself will come to see things this way once he gets to know the college and its ways better. And finally, he gently suggests that Charles is being overly sensitive – something to which (the Master goes on) Charles is understandably prone to being, given his working-class background. Charles leaves the conversation unmoved, continuing to believe that he has faced classist discrimination in the college, and dismissing the Master's testimony. Charles meanwhile is unaware

that some people from working class backgrounds (e.g. the Master) suffer from false consciousness, distorting their ability to recognise class-based oppression. (Srinivasan, 2018, 5-6)

DOMESTIC VIOLENCE Radha is a woman who lives in rural India. Her husband, Krishnan, regularly beats her. After the beatings, Krishnan often expresses regret for having had to beat her, but explains that it was Radha's fault for being insufficiently obedient or caring. Radha finds these beatings humiliating and guilt-inducing; she believes she has only herself to blame, and that she deserves to be beaten for her bad behaviour. After all, her parents, elders and friends agree that if she is being beaten it must be her fault, and no one she knows has ever offered a contrary opinion. Moreover, Radha has thoroughly reflected on the issue and concluded that, given the natural social roles of men and women, women deserve to be beaten by their husbands when they misbehave. (Srinivasan, 2018, 5-6)

Srinivasan asks us to agree that Nour's belief that her host is racist and Charles's belief that the college is classist are justified, while Radha's belief that a woman who does not do as her husband demands deserves to be beaten is unjustified. And she argues that only an externalist account of justification can deliver these judgments. Indeed, it seems that only a reliabilist account can deliver the first judgment, while a wider range of externalist views might accommodate the second.

Up to this point, Srinivasan's argument belongs to the first category enumerated above—it relies on our intuitive responses to a range of cases. But Srinivasan then extends the argument so that it comes to belong to the third category. She argues that, not only is the externalist best equipped to preserve our intuitions in the case of Nour, Charles, and Radha, but that an externalist concept of justification is best able to provide a certain sort of explanation that we need for political purposes. Let's see how.

Srinivasan notes that, when our intuitive responses to a vignette favour externalism over internalism, the vignette in question tends to be characterised by what she calls "bad ideology" (Srinivasan, 2018, 16). That is, in such vignettes, the protagonist lives in a society throughout which pervades a system of false beliefs that serve to sustain certain social oppressions, and the protagonist's relevant beliefs are significantly influenced by this system of false beliefs. Nour lives in a society with a racist ideology, Charles in a society with a classist one, and Radha lives in a misogynist society. Srinivasan also claims that, in such cases, we reach not for individual explanations, but for structural ones. That is, when we explain something about an individual's actions or thoughts, we are not satisfied by learning the features internal to the individual that gave rise to it. We want to know also the structural features of the society in which the individual is embed-

ded that gave rise to it. According to the intuitions that Srinivasan elicits in response to the case of Nour and Charles, both individuals are justified in their beliefs. Srinivasan thinks these facts call for structural explanation. She also thinks that, if our notion of justification is internalist, we cannot give such an explanation, while if it is externalist, we can.

[E]xternalism – in its insistence that justification can supervene on facts external to the agent’s own ken – is poised to vindicate what we might think of as a structural rather than merely individualistic notion of justification. A structural explanation gives an account of its explanandum by adverting to the larger system of which the explanandum is a part, rather than (solely) adverting to features of the explanandum itself. [...] Meanwhile, the internalist – in her insistence that justification supervenes on a subject’s mental states – is not poised to underwrite a structural notion of justification. Instead, she can only explain an agent’s justificatory status in terms that are intrinsic to the agent herself. Externalism, but not internalism, is poised to vindicate Trotsky’s claim that ‘escape from the web of the social lie’ is more than a matter of ‘mere individual effort’. (Srinivasan, 2018, 19)

It isn’t obvious to me that externalist accounts of justification are better placed to support structural explanations than internalist ones. Let’s consider the sort of straightforward reliabilism, for instance, that seems to account most naturally for the fact that Nour’s belief that her host is racist is justified. Nour has a reliable racism-detecting mechanism. She isn’t aware of how it works; perhaps she isn’t even aware that it’s reliable. But nonetheless it is. That’s the reliabilist’s explanation of the fact that Nour’s belief is justified—and it is individualistic, not structural. Now of course we can then ask for a structural explanation on top of that. And indeed we can give one, perhaps along the lines of Charles Mills’ explanation of why Black members of a White supremacist society are better able to pierce Trotsky’s ‘social lie’:

Often for their very survival, blacks have been forced to become lay anthropologists, studying the strange culture, customs, and mind-set of the ‘white tribe’ that has such frightening power over them, that in certain time periods can even determine their life or death on a whim. (Mills, 2007, 17-8)

We can extract from Mills’ account an explanation for why Black members of a White supremacist society have reliable belief-forming mechanisms concerning racism. It is a threat-detection mechanism, and it is clear what incentive a person who is the target of such racism has to develop one of those in a society in which such threats are common. So the reliabilist explains why someone’s belief is justified by pointing to the reliability of the



process by which it was formed; we can then supplement that individualistic explanation with a structural explanation of why the individual has the reliable belief-forming process. The two explanations neatly dovetail.

So Srinivasan is right to say that externalist accounts of justification are well-suited to fit into structural explanations. However, internalist explanations have the same feature. Let's return to Srinivasan's example of Charles. But now extend the example so that he later acquires the concept of false consciousness that he lacks in Srinivasan's version. At this later point, he thereby comes to understand why he is right to retain his belief that the college is classist and the Master is mistaken. So at this later point, his belief that his college is classist is internally justified. Now let's see what sort of explanation the internalist can give of this fact. As with the externalist, they begin with an individualist explanation: Charles's belief that his college is classist is justified because he can give an argument in its favour on the basis of his colleagues' behaviour, and he can appeal to false consciousness to explain why the Master's testimony does not undermine that argument. But the internalist can then go on to supplement this individualist explanation with a structural explanation, just as the reliabilist is able to give a structural explanation for why Nour has the reliable racism-detecting mechanism: Charles is able to give the justificatory argument for his belief and a further argument why the Master's testimony doesn't undermine that argument, because he lives in a society in which the notion of false consciousness has been identified and expounded, and the society is open enough that literature expounding and explaining this radical notion is available to Charles, and his situation permits him time to read this literature, absorb it, and reflect on it. That is, just as Nour's ability to pierce "the social lie" requires more than "mere individual effort", to use Trotsky's phrases, so does Charles's.

Indeed, compare the structural explanation we've just offered for Charles's justified belief and the paradigm example of a structural explanation that Srinivasan offers:

To explain that the dutiful housewife does the lion's share of the domestic labour because she prefers it that way is to give an individualistic explanation of her behaviour; to explain that the dutiful housewife does the lion's share of the domestic labour because that is what is socially expected of women is to give a structural explanation. (Srinivasan, 2018, 19)

Here, the individualist explanation gives the proximate cause of the housewife's behaviour, namely, that she has certain preferences, which are an internal state of her mind; the structural explanation then identifies the ultimate cause by providing an explanation of that proximate cause, namely, that she has those preferences because those are the preferences expected by society, they are taught to young women in this society throughout their

formative years, endlessly modelled for them in fiction and media, and there is severe censure in adulthood for those who do not have such preferences. The same is true in the explanation of Charles's belief in the extension of Srinivasan's classist college example that we just described. He believes that the college is classist because of the internal justification he can offer of it. That's the proximate cause, and it provides the individualistic explanation. But he is equipped to offer that internal justification only because of certain features of his society. Those features are the ultimate cause, and they provide the structural explanation.

In sum: it seems to me that externalism is no better able to accommodate structural explanations than internalism is. Both primarily give individual-level explanations: the externalist might appeal to the reliability of Nour's belief-forming mechanism to explain why her belief is justified; the internalist might appeal to Charles's ability to argue internally for his belief and to argue internally that he is right to ignore the Master's misleading testimony in our extended version of his case. But we can ask in both cases whether it's possible to give societal-level explanations into which those individual-level explanations fit. And in both cases, it is possible. As we quoted above, Srinivasan says that externalism is better able to provide structural explanations than internalism because of "its insistence that justification can supervene on facts external to the agent's own ken". But, while standard versions of externalism do indeed take external facts to partly determine whether an individual is justified, those facts tend to be pretty local to the individual, and not structural features of the society of which they are a part. Nour's belief that her host is racist is justified because the mechanism she used to form it is reliable; it tends to produce true beliefs. The fact that the mechanism is reliable is surely external, but it says very little about the whole society of which she is a part. Of course, to explain why she has the reliable belief-forming mechanism, we must advert to the society of which she is a part. But, similarly, to explain why Charles has the cognitive ingredients to provide an internal justification for his belief that his college is classist, we must again advert to the society of which he is a part. The two views seem to me similar in this regard.

So I'm reluctant to grant Srinivasan her claim that there is a political purpose—providing structural explanations for facts about justified beliefs—that externalism serves better than internalism. But even if we were to grant it, it doesn't follow that we should adopt an externalist concept of justification, and reject an internalist one. To draw that conclusion, we must make two further assumptions. First, we must be monists about justification. That is, we must assume that, at the end of our investigations, we should end up with a single true concept of justification, rather than two or perhaps more. Second, we must assume that there is no practical or political purpose that internalist notions serve better than externalist ones, and which we might place alongside the practical and political purposes that

Srinivasan takes to be played better by externalist notions than by internalist ones. I wish to argue that both assumptions are mistaken.

Upon reflection, it surprises me that so few epistemologists have heeded Alston's call for pluralism about the concept of justification (Alston, 2005). On this view, there are a number of concepts that have equally good claim to be our concept of justification and there is no concept that has a better claim than these. That is, the upshot of the lengthy debates about the nature of justification is that it doesn't have a single nature. 'Justification' is thus a polysemous term; the concept it refers to fragments into many different concepts. Some of those are internalist, some externalist: mentalism might be one, process reliabilism another, and a third might be a demanding hybrid view on which a belief must be reliably formed and must cohere with the subject's other internally accessible attitudes before it is justified. These different concepts might serve different ends. An internalist concept might serve our purposes when we want to decide whether or not to blame someone for something they did on the basis of their beliefs; a hybrid concept might be required for determining whether an individual has a right to non-interference in a particular situation; and so on.

Part of what is surprising about the lack of support for pluralism is that it seems the natural response when competent users of the concept—as I assume analytic epistemologists of the past fifty years to be—disagree so irreconcilably on certain of its basic features. And indeed philosophers have been quick to give this response in other situations. Suppose I take a coin out of my pocket and tell you it is a trick coin, biased either towards landing heads or biased towards landing tails. You say the coin is just as likely to land heads as to land tails. Are you right? Many will say yes, and many will say no. In this case, we resolve the standoff by saying that there are two concepts of probability in play. The first is epistemic, the other ontic. The first measures something like your degree of confidence, and on that concept, it is right to say that the coin is equally likely to land heads or tails, for my confidence in each outcome is the same. The second measures something about the world independent of our knowledge of it, and on that concept, it is not just as likely to land as tails. Indeed, it is either biased towards heads in which case it is more likely to land heads than tails, or it is biased towards tails in which case it is more likely to land that way than the other. These two versions of the concept of probability have been distinguished and even given their own names: 'credence' for the epistemic concept and 'chance' for its ontic cousin. So why not do likewise for the concept of justification and countenance internalist, externalist, and hybrid versions?

I think there are two plausible explanations—and neither excludes the other. On the first, proponents of each analysis of justification think that there are already terms that cover the concepts picked out by the alternative analyses on offer. For instance, Littlejohn is often quick to point out

that the concepts *reasonable* or *excusable* or *rational* often apply in those cases to which the internalist would like to apply the concept *justified*. And an internalist or a proponent of a hybrid view, might think that simply calling something 'reliably formed' suffices to cover the cases the reliabilist would like to capture. If this is right, there's no need to split the concept of justification. Better to use the concepts that we already have.

One problem with this is that the alternative concepts offered usually don't match up. For instance, most internalists typically don't consider everything that is excusable to be justified. If a belief is implanted in my brain without my realising, it is excusable, but not justified; similarly, if I calculate something quickly because I need to make a decision and I make a small error in my calculation, my resulting incorrect belief is excusable, but again not justified. Another problem is that different parties to the debate understand the various concepts in different ways. For instance, Stew Cohen (2016) is happy to equate rationality and justification, while Clayton Littlejohn (2012) is not. And I think Littlejohn is right. For those who count the boundedly rational among the rational, a belief formed on the basis of a base rate fallacy, for instance, might very well count it as rational because formed by a method that proves reliable in our evolutionary niche, but most will agree that it is not justified. So the other concepts are no less problematic than the concept of justification, and skirmishes at their borders are no more tractable.

Another explanation is that each party to the dispute thinks the concept of justification plays such a crucial role in our reasoning about certain important matters that it might be positively dangerous to hand over control of its use to one of their rivals. Suppose, for instance, you accept the following: if someone is justified in believing that the person approaching them means to harm them, then anyone who stops them defending themselves against this perceived impending harm is violating their right to non-interference. Then it's going to be extremely important to you to ensure that the concept of justification does not fall under the control of the internalists. The worry is that the concept of justification is embedded in many important practical or legal inferences. And if that is so, we must seek a single concept that makes these inferences valid.

This worry is really a practical one. It raises a concern not about the Alstonian pluralist view itself, but rather about the practical consequences were it to be accepted widely. But just as there is a practical problem here, so there is a practical solution. We need only ensure that those inferences that use the term in moral and legal thinking are amended to invoke not the ambiguous concept of justification, which contains many different precise concepts within it, but rather the precise concept that ensures the inferences go through. Even if you imagine that this solution could not practically be implemented, that is no mark against the philosophical position behind it. People are often still unable to distinguish clearly between epistemic and

ontic notions of probability, but that makes the distinction between them no less valid.

In sum: I don't think there are good reasons to resist pluralism about justification. This, then, opens the door to a view on which there are legitimate internalist, externalist, and hybrid versions of the concept of justification, perhaps answering to different intuitions, perhaps playing different theoretical roles in our epistemology, perhaps serving different practical, moral, or political ends.

Of course, you might say that what this really supports is not pluralism about justification, but nihilism.<sup>1</sup> In the end, I think these positions are reasonably close to one another. Of course, they seem to be as far apart as possible: according to pluralism, there are many concepts of justification; according to nihilism, there are none. But according to the nihilist, all the concepts that the pluralist takes to be concepts of justification are nonetheless legitimate concepts. They just don't count as concepts of justification. This might be because the nihilist thinks that the concept of justification was introduced in a particular way that precludes pluralism about the concept. For instance, they might claim that it was introduced as the unique good-making feature of beliefs with a particular property. Or it might be because the nihilist thinks that the various concepts that the pluralist counts as different concepts of justification are not sufficiently closely related to count as different versions of the same concept. Myself, I doubt that the concept of justification was introduced in a way that precludes pluralism: like most of our concepts, I doubt it was introduced with an explicit description attached at all. And I do think that the various concepts the pluralist countenances form a natural group: each identifies a particular good-making feature of beliefs, but none can claim to be the unique one. So, it seems to me, it is pluralism rather than nihilism that is the correct response to the arguments of this paper.

On the pluralist view, an internalist notion of justification need not serve a practical, moral, or political end at all in order to be included in our suite of justification concepts. And it certainly needn't serve a more important such end than the end served by an externalist notion. It might instead earn its keep by systematising a certain collection of robust intuitions, or by playing a much-needed role in our epistemological theorising. But I nonetheless want to argue that there is such a political end that internalist concepts serve.

First, we need to introduce the notion of an epistemic weapon, and to do that, we need the notion of an epistemic harm. So, what is an epistemic harm? We might think of it by analogy with other sorts of harm. We might say that I harm someone in the usual sense, if I do something that causes them to have lower all-things-considered utility—that is, less of what they

---

<sup>1</sup>Thanks to an anonymous referee for pushing me to consider this possibility.

value overall—than they would have had if I had refrained.<sup>2</sup> By analogy, we might say that I *epistemically* harm them if I do something that causes them to have lower *epistemic* utility than they would have had if I had refrained. Here, your epistemic utility measures how much you have of what you value, epistemically speaking. For instance, you might value knowledge, true belief, understanding, wisdom, evidence, or some combination of these things. So I might epistemically harm you if I block your inquiry, if I fail to tell you information you would like to have, if I hide evidence from you, if I intentionally deceive or mislead you, if I lie to you, and so on. We then say that an epistemic weapon is a means by which an agent—whether an individual, dominant group, or whole society—can cause epistemic harm to a target—whether an individual, an oppressed group, or a whole society.

Now, epistemic goods—such as knowledge, true belief, understanding, wisdom, and evidence—are unequally and unfairly distributed within our society. This is due partly to the inequities of our education systems, the prevalence of hermeneutic epistemic injustices, and unequal access to shared evidence, public debate, and the tools for individual theorising. But it is also due to the effects of other, more local epistemic weapons. A crucial part of a radical epistemological project is therefore to develop effective defences against those weapons. And, to do that, we must understand how different weapons work. But, as I will now argue, we need the internalist concept of justification in order to usefully categorise different sorts of epistemic weaponry in the service of this task. And thus that version of the concept plays an important political role, just as Srinivasan argues the externalist version does.

Consider Nour and the dinner party and her racist host; and consider Charles and his classist college and its apologist Master. Both are, in different ways, targets of the epistemic weapon of gaslighting. In both cases, they successfully defend against the weapon. But in both cases they are threatened by it. While it doesn't in fact inflict an epistemic harm, it could easily have done.

In both cases, what makes them potentially vulnerable to gaslighting is a hermeneutic injustice (Fricker, 2007). Or, perhaps better, it is simply that they lack an important concept or a body of evidence that would allow them to understand their experience in a way that would make them better

---

<sup>2</sup>This is close to the account of harm known in the ethics literature as the *comparative counterfactual account* (Feinberg, 1986; Klocksien, 2012; Hanna, 2015; Purves, 2019). It is probably the most popular account among ethicists, though it is taken to face two sorts of problem: (i) it entails that you harm someone merely by failing to benefit them; and (ii) it entails that you do not harm them if you inflict some injury upon them that prevents them from suffering a greater injury. Similar objections might be raised against the account of epistemic harm that I adopt here, but similar responses are also available. For a compelling response to (i), see (Purves, 2019); for (ii), see (Klocksien, 2012).

able to withstand gaslighting. Nour has the concept of racism, but she does not understand from the inside what are the features of behaviour that typically indicate the presence of that phenomenon, or perhaps she can access an inventory of those features, but does not understand why they indicate what they do. Charles is better equipped. He has the concept of classism, just as Nour has the concept of racism, but he also knows what indicates its presence, understands how it works, and so on. What he lacks is the concept of false consciousness and the empirical knowledge of how that usually manifests in society.

Nour's epistemic poverty makes her vulnerable to gaslighting herself. She has this gut feeling that her host was racist. But without understanding why, or whether the gut feeling is reliable, she might easily second guess herself and decide her judgment is not sufficiently internally justified to retain it. If she abandons this true belief on this basis, she'll suffer an epistemic harm, since a true belief is an epistemic good. The epistemic weapon will have done epistemic damage.

Charles's epistemic poverty is less severe than Nour's, and as a result he is largely safe from gaslighting himself. But it does make him vulnerable to being gaslighted by the Master of his college. He has the first-order evidence of his colleagues' comments and behaviour, which he takes to support his diagnosis of classism. But he also has the second-order evidence that the Master disagrees with his diagnosis. And further, he has second-order evidence that suggests that the Master is more reliable than him—the Master is also from a working class background, and he has more experience of the college and the context of his colleagues' behaviour. Based on this, he might feel his belief is no longer internally justified and abandon it. If he were to have the concept of false consciousness, he might apply it to the Master, and thereby explain away their disagreement and retain his belief that the college is classist based on his first-order evidence. But without it, the Master threatens to successfully, though unintentionally, gaslight him. If he does, he'll suffer an epistemic harm. Again, the epistemic weapon will have hit its mark.

In both Nour's cases and Charles's, then, the efficacy of the epistemic weapon relies on its targets responding to their evidence by forming or retaining only those beliefs that are internally justified. And gaslighting will always rely on this. That is part of what makes it so insidious. To be gaslit is usually to respond in an internally justified way to your total evidence. You have first-order evidence from which you can internally justify a particular belief; then you acquire second-order evidence that someone in at least as good an epistemic position as you are disagrees that the evidence supports your belief, or claims to have more evidence that undermines that belief; you then respond by thinking that you can no longer give an internal justification of your belief and therefore decide to abandon it.

Now consider a new case. In it, an unscrupulous prosecutor is present-

ing his case in a murder trial. Having found DNA at the murder scene, the police took DNA samples from all 100,000 people living in the town and tested each for a match. The first person they randomly selected was a match according to their test, and they arrested him and put him on trial. Now, this test gives 1% false positives and 1% false negatives. Thus, if the defendant had a prior probability of  $\frac{1}{100,000}$  of being guilty, then they only have a posterior probability of around  $\frac{1}{1,000}$  after incorporating the evidence of the match.<sup>3</sup> But the prosecutor knows that most jurors will neglect the base rate and conclude from this evidence that the defendant is 99% likely to be guilty. So he mentions the match between the defendant's DNA and the DNA found at the crime scene, and he mentions the false positive and false negative rates of the test used.

Again, this is an epistemic weapon. Here, it is deployed to ensure that the jury convicts. If they ignore the base rate and conclude that the defendant is 99% likely to be guilty, they might well take that to put the matter beyond reasonable doubt. Thus, in this case, the efficacy of the epistemic weapon relies on precisely the opposite response from the one on which gaslighting relies. Gaslighting relies on targets responding to their evidence in a way that is internally justified. This weapon, which exploits our tendency to commit the base rate fallacy, relies on them responding in a way that is internally unjustified.

There are many ways dimensions along which epistemic weapons differ. For instance, they differ in what they target. Some target your concepts, others your evidence, and others still target your beliefs directly.

If an epistemic weapon targets your beliefs, it might try to implant a false belief or it might try to remove a true one. For instance, I might try to implant a false belief by lying to you—this is perhaps the simplest epistemic weapon. Or I might draw your attention to a tempting but fallacious inference that would take you from one of your current true beliefs to a new false belief. Knowing that you are aware that a particular politician recently met with PR firm, I might say 'If she were trying to hide a scandal, she'd be sure to enlist image consultants', hoping that you'd reason by affirming the consequent and come to believe that she is trying to hide a scandal, which I know she is not. And someone might try to remove one of your true beliefs by overloading you with closely related information in

---

<sup>3</sup>After all, by Bayes' Theorem,

$$\begin{aligned}
 P(\text{Guilty} \mid +\text{ve}) &= \frac{P(+\text{ve} \mid \text{Guilty})P(\text{Guilty})}{P(+\text{ve} \mid \text{Guilty})P(\text{Guilty}) + P(+\text{ve} \mid \overline{\text{Guilty}})P(\overline{\text{Guilty}})} \\
 &= \frac{0.99 \times 0.00001}{0.99 \times 0.00001 + 0.01 \times 0.99999} \\
 &\approx 0.001
 \end{aligned}$$



a way that leads you to forget that true belief. So a murderer who wants a witness to forget what they saw him wearing on the morning of the killing might flood them with information about what he's been wearing every day for the past month.

Alternatively, an epistemic weapon might target your concepts. And it might do this by trying to impoverish your conceptual scheme or by trying to make the concepts you have ambiguous. To impoverish your conceptual scheme, someone might persuade you that one of your concepts is unusable and should therefore be abandoned. They might persuade you that the concept has vague boundaries, and also persuade you that reasoning with vague concepts is fallacious or very likely to lead you astray; or they might persuade you that it cannot be operationalised, and also persuade you that only concepts that can be operationalised are meaningful. To make one of your concepts ambiguous, I might flood you with what I claim are instances of that concept that are difficult for you to bring under a single coherent natural definition. Someone who, for political reasons, would like to make your concept of labour unusable, even if they don't wish to deprive you of it, might list off myriad different sorts of activity that they say count as labour, making you wonder whether applications of your concept are sufficiently determinate for you to use it in your political thinking and organising.

And finally, an epistemic weapon might target your evidence. To do this, it might try to keep information from you or it might try to overwhelm you with conflicting and complex information that swamps your cognitive capacity. In the former case, a politician might tell you that GDP has grown, but omit to mention that the poorest off are less well off than before. In the latter case, a climate change denier might show multiple complex graphs with little explanation in the hope that it will make the evidence seem more equivocal than it is.

So one way in which epistemic weapons might differ is that they target different parts of your epistemic state. But, as we saw above in the contrast between the case of gaslighting and the case of the unscrupulous prosecutor, they also differ in what makes an individual vulnerable to them. One weapon might rely on you forming a belief about something only if it is internally justified—gaslighting is an instance of this. Another might rely on you forming beliefs in a way that is internally unjustified—the prosecutor who exploits your tendency to neglect base rates is an instance of this. And it is easy to see that different policies will serve to defend groups and individuals against these different varieties: against those that rely on individuals forming or retaining only those beliefs that they can justify internally, we must look outwards to ensure that the evidence they obtain and on which their internal justifications are based is not misleading; against those that rely on individuals systematically forming beliefs in ways that are internally unjustified, we must look inwards and try to correct those

ways of coming to believe. The concept of internal justification, therefore, earns its stripes on the radical side in the political battle against epistemic weaponry; it helps us build stronger epistemic defences.

In conclusion: we should be pluralists about our concept of justification. 'Justification' is a polysemous term, and there are many precise concepts that fall under it. These different concepts play different roles. Some capture a particularly widespread set of intuitions. Some play an important theoretical role. Some play an important practical role by pinning down the concept we need when we are ascribing blame, or the one we need when determining whether a person's right to non-interference has been violated. And finally some further a radical political cause by providing structural explanations of how individuals who pierce through bad ideology can be justified in their beliefs, while others further that same cause by helping differentiate different sorts of epistemic weapon so we can better build defences against them.

## References

- Alston, W. (1988). An Internalist Externalism. *Synthese*, 74, 265–83.
- Alston, W. (2005). *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca, NY: Cornell University Press.
- Austin, J. L. (1956). A Plea for Excuses. *Proceedings of the Aristotelian Society*, 57, 1–30.
- Cohen, S. (2016). Theorizing about the epistemic. *Inquiry: An Interdisciplinary Journal of Philosophy*, 59(7-8), 839–857.
- Collins, P. H. (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. London & New York: Routledge, 2<sup>nd</sup> ed.
- Conee, E., & Feldman, R. (2001). Internalism Defended. *American Philosophical Quarterly*, 38(1), 1–18.
- Dotson, K. (2012). A Cautionary Tale: On Limiting Epistemic Oppression. *Frontiers: A Journal of Women Studies*, 33(1), 24–47.
- Dutant, J. (2015). The legend of the justified true belief analysis. *Philosophical Perspectives*, 29(1), 95–145.
- Feinberg, J. (1986). Wrongful life and the counterfactual element in harming. *Social Philosophy and Policy*, 4(1), 145–178.
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford, UK: Oxford University Press.

- Frye, M. (1983). Oppression. In *The Politics of Reality: Essays in Feminist Theory*. Trumansburg, NY: The Crossing Press.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23(6), 121–123.
- Goldman, A. (1979). What is justified belief? In G. S. Pappas (Ed.) *Justification and Knowledge*. Dordrecht: Reidel.
- Hanna, N. (2015). Harm: Omission, preemption, freedom. *Philosophy and Phenomenological Research*, 91(2), 1–23.
- Klocksien, J. (2012). A defense of the counterfactual comparative account of harm. *American Philosophical Quarterly*, 49, 285–300.
- Littlejohn, C. (2012). *Justification and the Truth-Connection*. Oxford, UK: Oxford University Press.
- Mills, C. W. (2007). White Ignorance. In S. Sullivan, & N. Tuana (Eds.) *Race and Epistemologies of Ignorance*, (pp. 11–38). State University of New York Press.
- Purves, D. (2019). Harming as making worse off. *Philosophical Studies*, 176, 2629–2656.
- Srinivasan, A. (2018). Radical Externalism. *Philosophical Review*.
- Williams, B. (1981). *Moral Luck*. Cambridge, UK: Cambridge University Press.