Abstract

When is a belief justified? I consider three sorts of arguments for different accounts of justification on the spectrum from extreme internalism to extreme externalism: arguments from intuitive responses to examples; arguments from the theoretical role of the term in epistemology; and arguments from the practical, moral, and political uses to which we wish to use the term. I focus particularly on the third sort, considering arguments from Clayton Littlejohn (2012) and Amia Srinivasan (2018) in favour of different versions of externalism. I offer counterarguments in the same vein for internalism. 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 part of that involved understanding when a belief counts as justified, since most 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. The debate 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 along a spectrum, with extreme internalism at one end and extreme externalism at the other. Where you lie on this spectrum is determined by the extent to which the conditions that suffice for and are required for justification

*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.*
concern matters internal to the subject who has the belief and the extent to which they concern matters external to them.

To illustrate this spectrum, let’s map out a few positions on it. First, the most extreme internalist position. According to this, whether or not you are justified depends only on your actual occurrent internal mental states, such as the perceptual experiences you are having at the time the justification of the belief is assessed, or memories that you are accessing at that time. Mental states cannot make a difference unless they are not only accessible to you but actually accessed at the time of evaluation. This is what Pappas (2017) calls actual access internalism, but for justification rather than knowledge. A slightly less extreme version permits mental states that are accessible whether or not they have been accessed, such as a memory that you could retrieve easily, even though you in fact haven’t. This is access internalism. And a less extreme version still permits mental states that are not even accessible to you, such as a feature you did not consciously notice in a scene you just observed, but which you nonetheless registered at a subconscious level. This is what Conee & Feldman (2001) call mentalism.

At the furthest externalist extreme is the claim—held by no-one, I think—that a belief is justified just in case it is true. According to this, providing your belief concerns only external matters, its status as justified or not is determined entirely by external matters. Slightly less extreme than this is process or indicator reliabilism (Goldman, 1979; Alston, 1988). Both appeal to an internal state: the belief-forming process, much of which is internal, or the grounds of the belief, which is often an internal state, such as a perceptual experience. But they consider mainly external features of those internal states: the proportion of the beliefs that the process produces that are true; the objective probability that the belief is true given that the subject has that ground for it.

Closer to the middle of the spectrum, we have various hybrid accounts. We might require, for instance, that the process by which you form your belief is reliable, but also that you are aware of that process, understand how it works to some extent, and your other beliefs support or at least do not undermine the view that the process is reliable. Such a position is neither purely internalist nor purely externalist.

In what follows, we’ll meet plenty of these positions. I am interested here in how we might choose between them. After all, as they are usually stated, they are rival accounts of a single concept of justification that many people will have prior to engaging in systematic thinking about epistemology.

The clearest inventory of arguments for and against internalism or externalism about justification 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 given by competent users of the concept 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 or political uses.

Before we meet these arguments, it’s worth noting that they each have a different target. The first sort of argument is 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. The second sort of argument is also concerned with that existing concept, but it is 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. The third is much less concerned with the existing concept, and more concerned with whatever concept might serve our practical 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 are the most useful practically.

Let’s turn now to the arguments.

In the first category: Consider Richard, who grows up in a society in which a false account of a particular period in its history is systematically taught in the education system and bolstered by the media. He believes this account. His belief is well supported by the evidence he has and there is no evidence available to him that he has ignored. From the internal point of view, his belief looks as good as it could; yet we intuitively think it is unjustified. Or so Alston (1989) says, and he argues on this basis that internalism gets this case wrong.

Or consider Nagi, who can very reliably detect fake Leonardo paintings, but does not realise she has this ability and does not understand any of the underlying mechanisms by which she does so. Standing in front of a painting that the National Gallery ascribes to Leonardo, she forms the belief that it is a fake. From the reliabilist point of view, her belief looks as good as it could; yet we intuitively think it is unjustified. If that’s right, at least the reasonably extreme reliabilist version of externalism must be false, though more moderate versions are not affected (BonJour, 1980).

Or consider Chester. They stand in an art gallery, looking at a painting that appears to them entirely red. They are told by the gallery curator that it is in fact a white canvas, but lit by red light. Nevertheless, Chester persists in their belief, and in fact they turn out to be correct. Though typically an honest individual, the gallery curator was lying. We typically think that Chester’s belief, while initially justified, becomes unjustified after the gallery curator lies to them. And it is often thought that this makes trouble for externalists, though the case is not clear cut (Lasonen-Aarnio, 2010).

Or consider Jane, who is a handless brain in a vat. Her experiences
are all the result of inputs from malicious scientists who wish to convince her that she lives in a world like ours, has hands, and so on. Like her handed counterpart in the actual world, Jane believes that she has hands. Intuitively, we feel that Jane’s belief is justified, perhaps because it seems that her evidence is the same as her actual world counterpart’s. But the externalist has a hard time accounting for this judgment. This is the New Evil Demon objection to externalism (Cohen, 1984).

In the second category: we might argue in favour of externalism by pointing out that only it can explain why beliefs formed in certain ways are justified while those formed in other ways are not (Goldman, 1979). What distinguishes beliefs formed on the basis of perceptual seemings and expert testimony, which are typically justified, from those formed on the basis of wishful thinking and implicit racial biases, which are typically not? Extern- al features of those belief-forming processes. Thus, we can only explain why we draw the distinctions we do if we accept some form of externalism.

In favour of internalism, on the other hand, we might appeal to a connection between our epistemic duties and justification. If I’ve done all that could reasonably be asked of me epistemically speaking, surely my belief is justified. But whether or not something can reasonably be asked of me depends only on states that are accessible to me. For instance, you cannot reasonably expect me to believe something if my accessible evidence makes it very unlikely, even if my inaccessible evidence supports it. So internalism must be true.

In this paper, I’m particularly interested in the third sort of argument. 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. Perhaps the most well known argument in this category 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 distinctly, 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 excus-ably 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 the next 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, in each case asks us to agree with her intuitive evaluation of a particular belief held by the protagonist in that story, and notes that only the externalist can match the intuitive verdict.

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”. 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 women deserve to be beaten by their husbands if they do not do as they command is unjustified (just as Alston took Richard’s belief in the account of the historical period to be unjustified above). 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 views might accommodate the second.

So far, so familiar. We are used to this argumentative strategy from our first category of argument. But, as Srinivasan points out, what is interesting about these cases is that they are structurally the same as cases that are traditionally used to argue for internalism. That is, for each, there is a vignette with what internalist and externalist will agree are the same epistemically relevant features, but to which we intuitively react with a judgment that only the internalist can match. The case of Nour is analogous to the case of Nagi above. Just as Nagi had an extremely reliable eye for a fake Leonardo, even though she doesn’t understand how it works or what features it detects, Nour can reliably detect racism, even though she doesn’t understand how or what features of the dinner party conversation led her to this conclusion. The case of Charles is analogous to the case of Chester above. Just as Chester has a reliable colour sense, but is tricked by the usually trustworthy gallery attendant, Charles has a reliable nose for classism, but is gaslighted by the Master of the college. And the case of Radha is analogous to the case of Jane above. Just as Jane is fed only misleading evidence about the existence of her hands, and has no access to evidence of her true situation, Radha is fed only misleading evidence about the morality of domestic violence, and has no access to evidence of the true moral situation.

Now, there are two ways to continue the argument from this point, and Srinivasan considers both. On the first, it is an argument that lives solely in the first category. Read thus, the externalist can use Srinivasan’s cases to bolster their position if they offer a compelling error theory for the intuitions we have in Nagi’s, Chester’s, and Jane’s cases, while the internalist can use this suite of cases to shore up their view if they can offer a compelling error theory in the cases of Nour, Charles, and Radha. Srinivasan considers one sort of error theory that the internalists might offer: they might argue that, in the cases of Nour and Charles, we are loath to say something negative about someone who has experienced racism or classism, and thus feel uncomfortable declaring their beliefs unjustified in a way that we don’t with Nagi or Chester; in the case of Radha, we are loath to say something positive about such a tragic case of internalised oppression, and thus feel uncomfortable declaring her belief justified in a way that we don’t with Jane. That is, our intuitions might be swayed by the moral weight of the situation in Srinivasan’s cases. Rightly, I think, Srinivasan rejects this error theory. But it’s not obvious this is the only error theory the internalist might offer.

What’s more, it’s not obvious that the error theory that Srinivasan offers on behalf of the externalist for our reactions to Nagi’s, Chester’s, and Jane’s cases really works. Let’s consider the first component of Srinivasan’s error theory first. She notes that externalist responses seem most plausible in cases characterised by what she calls ‘bad ideology’—that is, cases in which the protagonist lives in a society throughout which pervades a sys-
tem 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. And she attempts to explain why bad ideology cases provoke this reaction by saying that, in such cases, we seek structural explanations of phenomena, rather than individualistic ones, and externalism serves the former sort of explanation, while internalism serves the latter.

Now, it’s not obvious that a positively-valenced normative concept like justification is needed for the sort of explanation that Srinivasan seeks, nor even that it plays that explanatory role well. That explanation might be given better using a purely descriptive concept, such as the concept of a reliably-formed belief. In situations of bad ideology, individuals who occupy a particular social position can pierce through the social lie because their position bestows on them more reliable belief-forming processes concerning the nature of the society they inhabit.\(^1\) While the fact that individuals can pierce the social lie is a good thing, it is nonetheless just a fact that can be stated without normative language. And this suggests that we should seek to explain it without such language, just as it is a good thing that chloroquine kills the parasites that cause malaria, but we wouldn’t expect the explanation of that fact to include normative language.

Nonetheless, let’s grant Srinivasan her claim about such explanations. This does not explain why our internalist intuitions in the case of Jane, Chester, and Nagi are wrong, since they are not cases of bad ideology: while Jane, Chester, and Nagi might well live in societies in which there are pervasive systems of false beliefs propping up social oppressions, they don’t affect Nagi’s beliefs about the provenance of paintings, Chester’s about the colour of painting, nor Jane’s beliefs about her hands. Indeed, Srinivasan’s explanation in terms of bad ideology seems to neatly account for our different reactions to the two classes of cases without saying that one is right and the other wrong. That is, it seems to point not to externalism as the conclusion, but to one of the following: first, we might say that the concept of justification is disjunctive, so that in cases of bad ideology it is externalist while in cases of good ideology it is internalist; or, second, we might say that there are just two concepts here, so that there is an internalist one and an externalist one, and we are interested in the internalist one in cases of good ideology or where bad ideology has no causal effect, and we are interested in the externalist concept in cases of bad ideology. Nothing that Srinivasan says has ruled out these options.

Let’s turn now to the second component of Srinivasan’s error theory.

\(^1\)See, for instance, Charles Mills’ explanation of why Black people in a White supremacist society understand the racist nature of that society better than the White people in it. It contains no normative language. “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)
She claims that the cases of Nour, Charles, and Radha are commonplace; they are familiar and an all too prevalent part of our actual world. In contrast, the standard cases of clairvoyants (BonJour, 1980), deceptive artists (Lasonen-Aarnio, 2010), or brains in vats (Putnam, 1981)—from which the cases of Nagi, Chester, and Jane above are derived—are recherché. From this, she argues that our intuitions are likely to be more reliable in the former cases than in the latter. I have some sympathy with this error theory when it applies to the original versions of the cases that I have presented as Nagi and Chester; and I think it also casts doubt on the reliability of our intuitions in the case of Jane. But the versions I presented at the beginning seem reasonably commonplace.

Furthermore, even if Srinivasan were right, note that her point is only comparative. She thinks that our intuitions in ordinary cases are more reliable than in recherché cases. It doesn’t follow, however, that our intuitions in recherché cases are unreliable. Both might be reliable, but those in ordinary cases more so. The comparative claim only tells against internalism if we reject the pluralism about concepts of justification that I discuss and endorse below. Only if internalism and externalism are competing to be the single true concept of justification is the comparative claim a threat to internalism. If we permit many concepts of justification, and if we accept that the intuitions supporting internalism are reliable though less reliable than those supporting externalism, then we might admit both concepts.

Now, as Srinivasan points out, we might try to go further than this version of the first sort of argument, and instead establish externalism by moving to an argument in the third category based on the cases of Nour, Charles, and Radha. Perhaps we accept that the internalist makes no mistake in the case of Nagi, Chester, and Jane, while the externalist makes no error in the case of Nour, Charles, and Radha. So, from the point of view of our actual concept of justification, either it is disjunctive—sometimes internalist and sometimes externalist—or it is ambiguous between two concepts—one internalist, one externalist—and we use the same word in different circumstances to refer to them. But, the externalist might argue, if we appeal to the practical uses to which we wish to put the concept of justification, we can discern a reason to favour the externalist concept of justification over the internalist one. As Srinivasan suggests herself, correctly categorising Nour’s, Charles’s, and Radha’s beliefs is politically important, and more politically important, we might infer, than correctly categorising Nagi’s, Chester’s, and Jane’s.

I am happy to accept all of this. But it doesn’t follow that we should adopt an externalist concept of justification, and reject an internalist one. To conclude that 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: access internalism might be one, process reliabilism another, and a third might be a 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 an extreme reliabilist concept might serve our political end of giving structural explanations of how someone breaks through bad ideology to furnish themselves with true beliefs about their oppression.

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.
Indeed, it’s worth noting that philosophers have already divided the very concept of justification, though in a different way from the one that is proposed here. If I have evidence for a belief, but I’ve formed the belief not on the basis of that evidence but on the basis of wishful thinking, for instance, we say that my belief is propositionally justified but not doxastically justified. And we met the notion of personal justification above, when we were discussing Clayton Littlejohn’s view of excusable beliefs. So we already countenance different versions of justification. Why not also 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. On this 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 are such practical, moral, and political ends that internalist concepts play.

Here’s the first. We use an internalist concept of justification to signal that a person’s beliefs or actions are internally coherent, rationally comprehensible, based on foundations that they take to be correct, even if they’re in fact mistaken. And that is useful because it allows us to predict their future behaviour. If I see someone behave in a way that is incomprehensible to me, but learn that they are internally justified in doing so, I learn that they must have beliefs that support or rationalise that behaviour, and that allows me to predict how they will behave in other situations. One weakness of reliabilist concepts of justification, for instance, is that learning that someone’s belief is justified tells us little about their other cognitive states. It tells us only about the relationship between that belief and the part of the world that it concerns.

So there are pragmatic benefits to including the internalist concept of justification that go beyond its use in assigning blame. But these are not specifically political benefits to match those catalogued by Srinivasan for the externalist concept. To see that there are those sorts of benefits too, we
need look no further than the sort of case that Srinivasan herself considers.

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 exactly by analogy with our account of other sorts of harm. Thus, just as 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 value overall—than they would have had if I had refrained, so we might say that I epistemically harm them if I do something that causes them to have lower epistemic utility—that is, less of what they value, epistemically speaking—than they would have had if I had refrained. 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 concept does.

Consider Nour and her racist host, and 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—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 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 in some ways. He has the concept of classism, 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, 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 gaslighting by the Master of his college. Without the concept of false consciousness, and with the evidence that the Master takes there to be no classism, he might feel his belief is no longer internally justified and abandon it. If he does, he’ll suffer an epistemic harm. Again, the epistemic weapon will have hit its mark. Thus, in both cases, the efficacy of the epistemic weapon relies on the targets responding to their evidence by forming or retaining only those beliefs that are internally justified.

Now consider a new case. In it, an unscrupulous prosecutor is presenting 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.\(^2\) 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\(^2\)After all, by Bayes' Theorem,

\[
P(\text{Guilty} | +ve) = \frac{P(+ve | \text{Guilty})P(\text{Guilty})}{P(+ve | \text{Guilty})P(\text{Guilty}) + P(+ve | \text{Guilty})P(\text{Guilty})} = \frac{0.99 \times 0.00001}{0.99 \times 0.00001 + 0.01 \times 0.99999} \approx 0.001
\]
our tendency to commit the base rate fallacy, relies on them responding in a way that is internally unjustified.

There are many dimensions along which epistemic weapons might differ. Does it target your concepts, or your evidence, or your beliefs? If it’s your beliefs, does it try to implant a false belief or does it try to remove a true one? If it’s your concepts, does it try to impoverish your concepts or make the ones you have ambiguous? If it’s your evidence, does it try to keep information from you or overwhelm you with conflicting and complex information that swamps your cognitive capacity? But they also differ in what makes an individual vulnerable to them: Do they form beliefs about the relevant subject matter in a way that is internally justified or not? And it is easy to see that different policies will serve to defend groups and individuals against these different varieties. The concept of internal justification, therefore, earns its stripes on the radical side in the political battle against epistemic weaponry.

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, such as when a mentalist concept covers the cases of Nagi, Chester, and Jane above; or an externalist concept captures the case of Richard, Nour, Charles, or Radha. Some play an important theoretical role, such as when internalism accounts for epistemic blame, or externalism explains why certain foundational beliefs, such as those formed on the basis of experience, memory, or testimony are justified. Some play a 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 can pierce through bad ideology, while others further that same cause by helping differentiate different sorts of epistemic weapon so we can better build defences against them.

References


