Erkenntnis
https://doi.org/10.1007/s10670-023-00778-0

ORIGINAL RESEARCH

Internalism and Culpable Irrationality

Karl Gustav Bergman[1,2]

Received: 10 April 2023 / Accepted: 10 December 2023

Abstract
According to internalism about rationality, the ir/rationality of a subject depends only on how things appear from her subjective perspective. According to culpabilism, rationality is a normative standard such that violations of rationality are (at least sometimes) blameworthy. According to a classical line of reasoning, culpabilism entails internalism. I argue that, to the contrary, culpabilism entails that internalism is false. The internalist cannot accommodate the possibility of culpable irrationality.

1 Introduction

Many have argued or taken for granted that the rationality or irrationality of a subject's beliefs depends only on how things seem to her from her own internal, subjective perspective (call this view *internalism*). Many have also supposed that rationality is *normative* in a qualified sense, namely, that subjects are blameworthy or culpable for (at least some of) their violations of the requirements of rationality (call this view *culpabilism*). And some have seen, in culpabilism, an argument for internalism: *because* violations of rationality are culpable, subjects must be able to ascertain, from their own internal perspective, whether they are in conformity with the requirements of rationality.

Both views can be and have been challenged, and so has the purported inference from the one to the other. In this paper, I will go one step further. I will argue that, given culpabilism, there is good reason to think that internalism is *false*. Internalism, I will argue, leaves no room for culpable violations of rationality.

✉ Karl Gustav Bergman
karl.bergman@filosofi.uu.se

[1] Uppsala University, Box 627, 751 26 Uppsala, Sweden

[2] Universitat de Barcelona, Montalegre 6, 08001 Barcelona, Spain

Published online: 02 February 2024

It will be possible to read my argument in the contrapositive direction, as an argument from internalism to the falsehood of culpabilism (cf. Glüer & Wikforss, 2009). However, a number of developments in recent and not-so-recent philosophy conspire to put independent pressure on internalism.[1] In this dialectical setting, an argument for the incompatibility of internalism with culpabilism serves to further weaken internalism.

I proceed as follows. In Sect. 2, I introduce internalism. In Sect. 3, I introduce culpabilism and explain why culpabilism may appear to support internalism. In Sects. 4 and 5, I offer an argument from culpabilism against internalism. Section 6 concludes.

2 Internalism

In the introduction, I defined internalism as the view that "the rationality or irrationality of a subject's beliefs depends only on how things seem to her from her own internal, subjective perspective."

As thus defined, internalism has three features:

1) It is a claim about the rationality and irrationality…
2) …of *beliefs*….
3) …to the effect that they depend only on how things seem to the subject from her own internal, subjective perspective.

All three require some commentary. I will discuss them in turn.

Regarding (1): internalism is a claim about *rationality* (and irrationality). "Rationality" is a slippery term, and not everyone agrees that it suffices to unambiguously identify a determinate phenomenon. I shall have more to say on how I purport to identify the target phenomenon towards the end of this section. For the moment, I must trust in the reader's existing grasp of the notion.

Regarding (2): the restriction to *beliefs* is a matter of economy of presentation. There are analogous internalist claims about the rationality of other kinds of attitudes, like desires or intentions. These may or may not be vulnerable to arguments analogous to the one I give here. I believe they are, but that is not a claim I propose to defend here.

In fact, the view I will discuss is more restricted than the provisional definition suggests, because it concerns the rationality or lack thereof, not of individual beliefs, but of *groups* of beliefs. An individual belief is presumably rational or irrational only relative to some context. This context can include the subject's other beliefs as well as her non-belief mental states and possibly (though this, of course, is part of what's at stake in the debate on internalism) non-mental features of the world. When a belief is rational, it stands in the right relations to these contextual features, whatever those relations are. I will focus on the relations that a belief must have to

[1] E.g. (Alston 1986; Millikan 1993; Sorensen 1998; Williamson 2000; Srinivasan 2015; Wedgwood 2017, 166 ff.).

the subject's *other beliefs* in order to qualify as rational. An equally good way of talking about the same thing is to talk in terms of the relations that must hold *among the subject's beliefs* for them to qualify as rational. This latter idiom is the one I will employ below.

Regarding (3): The formulation in terms of "how things seem from a subject's internal, subjective perspective" is intended to capture a family of views according to which the ir/rationality of a subject's beliefs depend only on facts that share a certain *epistemic* property—the property of being apparent to or accessible to the subject's own internal, subjective perspective. I will call this epistemic property "transparency" and say that insofar as some fact possesses it, that fact is "transparent" or "transparently accessible." I will assume that transparency implies infallibility: if a fact is transparent, and the subject attempts to determine whether the fact obtains, they cannot fail to attain the right answer. More detailed discussion of the notion of transparency will follow, as the details become relevant.

Internalism construed in these epistemic terms is not the only view defended in the literature under the name "internalism". Ralph Wedgwood, for instance, defends a version of internalism where the internality of the facts that determine the rationality of a set of attitudes is not a matter of their epistemic accessibility but of their ability to directly influence mental causation (Wedgwood, 2017, chap. 7). The distinction drawn by Wedgwood evokes a closely related debate within the literature on epistemological justification between "accessibilist" versus "mentalist" conceptions of internalism, going back at least to (Feldman & Conee, 2001). The current paper, however, will focus on the epistemic version of internalism and will, accordingly, reserve the term "internalism" for that view.

To get a better sense for internalism, let us consider one way in which it could turn out to be false. Suppose that the following is a requirement of rationality, as may seem plausible:

> **No direct contradictions.** The rationality of a subject's beliefs requires that she does not at the same time believe a proposition P and its direct negation $\sim P$.

Given this supposition, and given internalism, it must be transparent to the subject whether or not her beliefs directly contradict each other. If, for example, a subject believes both that *London is pretty* and that *London is not pretty*, she must be able to tell, purely through introspection, that those beliefs are contradictory.

This example, incidentally, serves to illustrate some of the wider ramifications of the present issue. Some philosophers have argued that *semantic externalism*—an otherwise deeply influential view in the philosophy of language and mind—entails that contradictions among beliefs are, at least sometimes, nontransparent (Boghossian, 1994; Millikan, 1993). Consider this version of a classic case due to Saul Kripke (1979, 254–55):

> **Pierre.** Pierre is a Frenchman who has grown up in France where he has learnt about London through hearsay. Among the things people have told him about London is that (in Pierre's native French) "*Londres est jolie*." On the basis of this testimony, he comes to believe about London that it is pretty. Later, Pierre

moves to London and settles in an ugly part of the city. Based on his own eyewitness experience, he comes to believe about London that it is not pretty. Crucially, Pierre is never informed of the fact that he has moved to the same city he once heard about under the French name "*Londres*," and so he retains his old belief about London, that it is pretty.

Pierre has two beliefs about London, that it is pretty and that it is not pretty, but it is not transparent to Pierre that his beliefs are about the same city. Such failure of transparency of coreferentiality, or *de re* aboutness, is a quotidian phenomenon to post-Fregean philosophers, and need not, on its own, pose a problem for internalism (about rationality). There's a problem only if we also think, as Boghossian and Millikan do, that semantic externalism entails that these two beliefs—in virtue of attributing contradictory properties to the same city—are contradictory. If so, and if *No direct contradictions* is true, then semantic externalism entails the falsehood of internalism about rationality, because these two beliefs would then be directly contradictory, in violation of rationality, but this fact would not be transparent to Pierre.

If the above argument is sound, we are faced with a trilemma: We must either abandon internalism about rationality, abandon semantic externalism, or abandon the view that *No direct contradictions* is a requirement of rationality. None of these options seems very attractive.[2]

Why, then, should we believe in internalism? The most immediate case for the view rests on pure intuition. Consider the following case:

Internal twins. Two subjects, A and B, are exactly identical with regard to their internal, subjective perspective. Internally, things seem the same way to A as they do to B. But whereas A lives in our world, B lives in an evil-demon-world where all their experiences are produced by an evil demon and all their beliefs therefore false. (Adapted from Wedgwood, 2017, 161)

Despite major differences with regard, both to external situation and to externalistic features of their mental states such as their truth-value, A and B intuitively seem to be alike with respect to rationality: B's belief, intuitively, are rational to the same degree that A's are. This suggests that internalism is true: ir/rationality depends only on features of the subjects' internal perspectives.

If we reject internalism, on the other hand, we seem committed to the counter-intuitive view that A and B could differ in rationality, given only the right variation in their external circumstances. However, I don't believe this speaks conclusively in favor of internalism. For one, intuitions can be challenged, leading to dialectical

[2] Against the background of this larger dialectic, the present paper can be seen as a call for choosing the trilemma's first horn. But the argument against internalism to be offered below is intended to be independent of particular views on what rationality requires. It may be, for instance, that *No direct contradictions* is not a requirement of rationality—indeed, that rationality makes no requirements at all that constrain the *logical* relations among beliefs, defined over their propositional *contents*. If this were so, there would remain no dialectical pressure from semantic externalism against internalism about rationality, but the argument below would still be a challenge to the latter view.

standstill.[3] Such challenges to the reliability of intuitions carry particular force when the notion under investigation is already proto-theoretical, like rationality.

For the same reason, and as already mentioned, we cannot take for granted that "rationality" names a single unified property. If there are several properties tracked by this term, we may even grant to the internalist that one of these may be such that A and B possess it to the same degree. This concession would amount to little. The internalist would still need to show that the property in question is also capable of playing any of the *other* roles that we associate with the label "rationality". If not, then the admission that such a property exists and is among the things that can legitimately be called "rationality" will have little significance for any of the philosophical questions about rationality we are interested in or any of the uses to which we want to put the notion. There would be no guarantee that when we ask any of the interesting questions about rationality—whether, say, belief in conspiracy theories is rational, or whether the Wason selection task is proof of pervasive human irrationality—that what we're asking about, or *should* be asking about, is the internalistic property.[4]

For that reason, if the internalist intends her view to be of any wider theoretical interest, she should, at a minimum, offer some argument to the effect that internalism is the correct view about a property that is also capable of playing some of the other theoretical roles that we expect rationality to play.

One initial suggestion is that denying internalism amounts to, in Boghossian's words, "blur[ring] the distinction between errors of reasoning and errors of fact" (Boghossian, 2011, 458). One role that rationality is supposed to play is that of constituting an ideal or standard of right and proper reasoning. If internalism is false, however, then whether or not a subject succeeds in meeting the standard of rationality will not be entirely up to her own ratiocinative powers, but will, as it were, require the cooperation of the epistemic environment. Errors of reasoning will turn out to also sometimes be errors of fact, and Boghossian's distinction will be blurred.

But the insistence on a sharp distinction between reasoning and fact is bound to strike the skeptic as dogmatic, even as little more than a restatement of internalism itself. Are there independent grounds for insisting on the distinction?

This is where *culpabilism* comes into play, the view that rationality is supposed to play the role, not only of a standard of proper reasoning, but of a *normative* standard, failure to conform with which can render the subject culpable. If so, and if the internalist can show that rationality, in order to be a normative standard of this kind, must be internalist, then she has successfully met our challenge. Next, we turn to consider the prospects for this argument.

[3] Not everyone shares the internalistic intuitions that cases like *Internal twins* are supposed to elicit. (Millikan 1993, 348) is a prominent example to the contrary.

[4] In recent decades, the philosophical methodology of intuitions and cases have come under more general attack (Machery 2017; Millikan 2001; Stich 1993). This criticism, with which the present author finds himself broadly in sympathy, is also reason to be cautious of relying too heavily on said methodology—in addition to the reasons deriving from the slippery character of the term "rationality" specifically. This is especially the case since at least some of this general skepticism is motivated by the sort of semantic externalism that, as we saw above, generates dialectical pressure against internalism about rationality. The two sets of issues, then, are not dialectically independent.

3 Internalism from Culpabilism

Culpabilism is the view that rationality constitutes a standard or norm to which people have a *responsibility* to live up and to which they can, accordingly, be held *culpable* for failing to live up.

Culpability attaches in the first instance to acts, either in virtue of those acts themselves or in virtue of their outcomes. Since our topic is the ir/rationality of beliefs, specifically, the relevant acts will typically be acts of belief-formation, -revision, or -retention (the latter type of act is perhaps better thought of as an *abstention* than an act: an abstention from revising one's beliefs). From this point on, to avoid clutter, I'll talk about belief-*formation* and take the "or -revision or -retention" as read.

According to culpabilism, when acts of belief-formation violate the requirements of rationality or result in belief-combinations that violate those requirements, the subject is culpable or is at least *pro tanto* culpable or a candidate for culpability. For reasons to be discussed below, it would be too strong a view that held every violation of rationality to be culpable without admitting the possibility of exonerating circumstances; but culpabilism entails that at least *some* violations of rationality are culpable; namely, those performed in the absence of exonerating circumstances.

Is culpabilism true? Not everyone agrees that rationality is a standard of responsibility as culpabilism understands it. There is a longstanding and ongoing debate on whether, and in what sense, rationality is "normative". Most of this debate concerns whether there are *reasons* to be rational (Broome, 2007; Kolodny, 2005; Lord, 2018) or whether we *ought* to be rational (Wedgwood, 2017). Presumably, if we have no reasons to be rational (as Kolodny argues) or if it's not the case that we ought to be rational, then we are not culpable for violating rationality either (it is less clear whether the implication holds in the opposite direction).

Others (Glüer & Wikforss, 2009, 2013) have argued that standards of rationality cannot be *guiding* norms in the sense of norms that agents use to guide their actions. Again, presumably, if standards of rationality are not guiding, we cannot be held culpable for violating them.

Nevertheless, the idea that rationality is a standard of responsibility in this sense have appealed to many writers.[5] Here, we need not defend it, since we are only interested in what follows *if* it is true. Let us, then, proceed directly to examining the basis for the idea that culpabilism presupposes internalism.

In Sect. 2, I characterized transparency as the "property of being apparent to or accessible to the subject's own internal, subjective perspective," and added that this entails infallibility: if a fact is transparent, and the subject attempts to determine whether the fact obtains, she is guaranteed to succeed. This latter feature underwrites a further key feature of transparent facts: in trying to find them out, the subject is in an important sense *in control*. Specifically, she does not require the cooperation of sometimes recalcitrant external circumstances. Accordingly, *failures* to ascertain such facts are, in a correlative sense, entirely *due to* the subject.

[5] See, for instance, (Bonjour 1980, 59; Boghossian 1994, 42, 49; Brown 2004, 190–91; Kiesewetter 2017, 29; Lord 2018, 4). According to one commentator (Boult 2023), the conception of belief as subject to culpability has enjoyed a recent uptick in popularity.

On the other hand, in trying to ascertain some *non*-transparent fact, the subject's success or failure will be partly due to the cooperation or lack thereof of external factors, factors over which the subject lacks control. To illustrate this, let's return to the case of *Pierre* from Sect. 2. Pierre, recall, believes two incompatible things about London—that it is pretty and that it is not pretty. The coreferentiality of his beliefs is not, however, transparent to him. If he were to attempt to find out whether these beliefs concerned the same city, then try as he might, he could still fail. The world could fail to be forthcoming with the information that Pierre would need in order to realize that the two beliefs concerned the same city.

Enter now a very widespread and plausible idea about the conditions for responsible agency: that an agent is responsible only for what is due to her. Here is the idea expressed by Thomas Nagel:

> [I]t is intuitively plausible that people cannot be morally assessed for what is […] due to factors beyond their control. (Nagel, 1979, 138)

Nagel is here talking about *moral* assessment, and he does so in the context of a discussion of moral luck, the purported phenomenon whereby somebody can be rendered morally assessable and responsible for actions and outcomes that were partly due to external, coincidental factors. The quote gives voice to the intuitive supposition that moral luck cannot exist. But the intuition does not seem restricted to the moral domain, narrowly construed: it is an intuition about the conditions, not for *moral* responsibility, but for responsibility per se.[6]

What does it mean for something to be "due to" a subject? One common articulation is in terms of *voluntary control:* For an act to be due to a subject, the act must have been *willed* by the subject. But if this is the case, it seems our culpabilist, who holds that violations of requirements on rational *belief* are sometimes culpable, is committed to the controversial doctrine of *doxastic voluntarism*. Alternatively, the culpabilist can defend some weaker conception of what "due to" means in the case of belief-formation acts. Perhaps it means that the belief is formed by a mental faculty that is responsive to epistemic reasons (views in the vicinity are defended by Owens, 2000; Ryan, 2003; Steup, 2012).

We shall have reason to return to these delicate issues below, but for present purposes, it will suffice with an intuitive grasp of the notion of "due to" as drawing a distinction between that which has its origin in, on the one hand, the subject herself (as it were *qua* subject), and on the other, external vicissitudes.

With these ideas in hand, we can begin to see how the argument from culpabilism to internalism is supposed to go. In summary, the idea is this: in order for a subject to be culpable for a violation, the violation must be due to the subject herself. Internalism stands as a guarantee that when a subject violates rationality, the violation will be entirely due to her. Thus internalism stands as a guarantor of culpable irrationality in a way that rival views cannot. Thus culpabilism entails internalism.[7]

[6] The parallel between the problems of moral luck and rational or "logical" luck is drawn explicitly by Roy Sorensen (1998).

[7] Arguments that are essentially versions of this one are discussed by (Alston 1986) and (Wedgwood 2017, 167).

This rough sketch captures the intuitive core of the argument from culpabilism to internalism, but it could stand some refinement. As we refine it, we shall see it gradually fall apart. The exercise will nevertheless be informative.

Here is a first attempt at a slightly more formal version:

The argument from culpabilism to internalism.

1. Subjects are at least sometimes culpable for violating rationality. (Premise).
2. If a subject is culpable for violating rationality, the violation was due to her. (Premise).
3. Unless internalism is true, a subject's violations of rationality are never due to her. (Premise).
4. Unless internalism is true, subjects are never culpable for violating rationality. (From 2, 3).
5. Internalism is true. (From 1, 4).

The first premise follows directly from culpabilism, and premise (2) is the condition on responsibility already discussed.

What about premise (3)? As we will see, this is the argument's weakest links, but let us begin by attempting to reconstruct the reasoning that might lead someone to endorse it. A promising starting-point is the below principle, which gives shape to an idea already informally sketched above:

> **Due-to requires transparency.** Unless it is transparent to a subject which acts of her would constitute, or lead to, violations of rationality, then whether or not she avoids those violations will be contingent on outside factors and will therefore not be due to her.

Getting from *Due-to requires transparency* to (3) is no trivial matter, however. (3) says that unless internalism is true, a subject's violations of rationality are *never* due to her. But internalism is the view that the determinants of ir/rationality consist entirely of facts that are transparent to the subject. The negation of internalism is therefore consistent with it *sometimes* being transparent to the subject when an act of her would constitute a violation of rationality. Thus, for all that *Due-to requires transparency* says, the negation of internalism might still be consistent with the possibility of culpable irrationality and hence with culpabilism.

What *Due-to requires transparency* shows, if anything, is that unless internalism is true, subjects will *sometimes* be non-culpable for their violations of rationality. But this is of little help to the internalist unless she is prepared to endorse the view that violations of rationality are *always* culpable, a view we can call "strong culpabilism." Regrettably for the internalist, strong culpabilism is not very plausible. Consider: in the *moral* case, there are all sorts of ameliorating circumstances that can exculpate a would-be moral transgressor. Why should we expect the case of rationality to be any different? Even if irrationality is *pro tanto* culpable, it would be difficult for the transparentist to maintain that *no* irrational belief admits of exculpation (cf. Wedgwood, 2017, 168).

Putting aside these difficulties with inferring (3) from *Due-to requires transparency*, is the latter principle itself plausible? Is transparency really required for a violation to be due to a subject? If a subject performs an act that happens to violate some norm, then even if the subject was not privy to the fact that the act did violate the norm, there is an obvious sense in which the act, and hence the violation, was "due to" the subject—it was she, after all, who performed the violating act.

A more demanding sense of "due to" must be in play, if *Due-to requires transparency* is to be plausible. Something like this may do the trick: a subject's norm-violation fails to be due to her, in the more demanding sense, if the violation could occur despite the subject's best intentions to the contrary. This would be the case in the absence of a *reliable connection* between the subject's ambition to conform to the norm and her actual success in this endeavor. Non-transparency will entail such an absence, as the case of Pierre illustrates. Try as he might, Pierre cannot ensure that he doesn't believe incompatible things of the same city just by willing it to be so. The epistemic environment has to cooperate, supplying the empirical information that Pierre would need to guide his belief-forming in conformity with the norm. According to the present proposal, Pierre would therefore not be culpable for his troublesome beliefs, just as the intuitive verdict would have it.[8]

On this conception of culpability, a violation could only be culpable if the subject would have been guaranteed to avoid it, if only she had tried. Thus, a violation is culpable only if the subject either a) acted in reckless disregard of the norm or b) actively sought to violate it.

What has now emerged, however, is an implausibly strong set of requirements on culpability for irrational beliefs. To begin with, we are often held responsible for transgressions whose transgressive status we could have learned only through nontransparent means. The traffic rules of a foreign land can be ascertained only through empirical means, and the same is true for the fact about which side of the road one is currently driving on, yet if I went driving in the UK and drove on the right-hand side of the road, I could—in many cases—reasonably be held culpable for my violation of the local traffic rules.

I would be culpable if I, though informed (by empirical means) of those rules and able to determine which side I was in fact driving on, still opted to drive on the wrong side of the road (due to malice or a death wish or whatever). I would also, in many cases, be culpable even if I were *not* informed of those facts. A subject can presumably be culpable for violating a norm unknowingly, as long as her ignorance was *itself* culpable. What makes ignorance culpable is itself a contested matter, but one paradigmatic case is when the ignorance derives from a previous

[8] Something like this conception of culpability might lie behind Boghossian's suggestion that "[a] thinker is to be absolved for believing a contradiction, provided that the contradictory character of the proposition he believes is inaccessible to mere a priori reflection on his part" (Boghossian 1994, 49). Boghossian goes on to complain that "against the background of a non-transparent conception of propositional content, any contradictory proposition will satisfy that description … practically any contradictory belief will be absolvable under the terms of this proposal" (ibid.)—in other words: without transparency, there's no room for culpable violations of rationality.

A very similar view of the conditions for being—not culpable, but unjustified—underlies Carl Ginet's classic argument for internalism about epistemic justification (Ginet 1975, 28).

act of omission or neglect to take available routes of inquiry, where this act itself meets reasonable standards for culpability. This will obtain in many versions of the traffic example. Though empirical, the means to learn the British traffic rules are, in most normal circumstances, readily available to me, and I could reasonably have been expected to avail myself of those means before venturing out on the British roads.

The same can be said about our friend Pierre. Under most realistic elaborations of the scenario, Pierre could easily have looked up "*Londres*" and "London" in, respectively, a French and English atlas; or asked his friends and relatives about the matter; or similar. And it seems that such epistemic due diligence is the least that can be expected of someone about to move to a new country. Given these elaborations, it wouldn't be far-fetched to hold Pierre culpable for his troublesome beliefs. He should have known better (cf. Faria, 2009).

In these cases, subjects are seemingly culpable despite not having transparent access to the facts of their violations and, sometimes, despite actually being ignorant of those violations. The internalist now seems committed to denying these intuitions.

The argument from culpabilism to internalism, then, looks rather weak. Naturally, there's further moves the internalist could make. We shall not tarry with these, however. If the argument I'm about to make is sound, there will be no inference from culpabilism to internalism, because culpabilism will turn out to imply that internalism is false. This, in turn, is because internalism simply *cannot accommodate the possibility* of culpable irrationality.

My argument for the last claim, in outline, is as follows. As far as I can tell, there are two ways in which a subject could end up culpable for a violation of rationality:

1. The subject performs the violating act *without* knowing that it violates rationality. She is culpable for her ignorance and thus for the violation.
2. The subject performs the violating act knowing full well that it violates rationality, and she is culpable for that violation.

To accommodate culpable irrationality, internalism must accommodate the possibility of either (1) or (2) (or both). I will argue that it cannot. It cannot accommodate (1) because it cannot accommodate the possibility of culpable ignorance of a violation of rationality, and it cannot accommodate (2) because nobody will ever *deliberately* end up wantonly violating the requirements of rationality.

In the following two sections, I spell out my arguments for these claims, in order.

4 Ignorance and Neglect

We saw above that a subject can arguably be culpable for a violation even if she is ignorant of its status as a violation. Most authors agree that this requires that the ignorance is itself culpable, which in turn requires that the episode or omission that brought about the ignorance meet relevant standards. Holly Smith (1983) coins the evocative term "benighting act" for such an ignorance-generating episode. Not all authors agree, however, that the episode in question has to be a deliberate act (cf. Clarke, 2014; 2017; Rudy-Hiller, 2017); so a more theoretically neutral terminological choice would be "benighting event." Let us say, then, that a subject can be culpable for a violation performed due to ignorance, provided her ignorance is due to a benighting event for which she is in turn culpable.[9] On the face of it, internalism would seem to have trouble accommodating the possibility of this sort of ignorance-based culpable violation of rationality—for the simple reason that it would seem to have trouble accommodating the existence of the relevant benighting events in the first place. The whole point of internalism is that it is *transparent* to subjects whether their beliefs are in violation of the requirements of rationality. If so, it is hard to see how a subject who violates rationality could ever fail to be ascertained of that violation. If she couldn't, there is no way for her to be ignorant, hence no way for her to be culpably ignorant.

There are two ways to resist this simple point.

1. Appeal to *introspective effort*.
2. Appeal to *higher-order ignorance*.

I'll discuss these in turn.

4.1 Introspective Effort

In Sect. 2, I defined transparency as "being apparent to or accessible to the subject's own internal, subjective perspective." I deliberately made this definition open-ended, in the interest of casting a fairly large net. It is now open to the internalist to claim that transparency as thus defined can accommodate the following picture: though the ir/rationality of my beliefs is indeed transparently accessible to me, it is not, as it were, *directly given*, self-intimating or "luminous" (Williamson, 2000, chap. 4). In other words, the access is not automatic: I need to deliberately engage in some kind of cognitive activity, exert some kind of introspective effort, to *get* access. This leaves open the possibility that I might fail to exert this effort (maybe I neglect to do it because I'm too lazy or too busy, or I simply forget) and so fail to foresee the irrationality that results from my act. If this failure to exert the needed effort is

[9] Some writers, like Randolph Clarke (2014; 2017), deny that the ignorance that precipitates a culpable act itself must be culpable, though he agrees with the majority that there are some standards it must meet. Whether culpability-precipitating ignorance is itself strictly speaking culpable is immaterial to my argument, so the reader can take the phrase "culpable ignorance" in what follows as short for "ignorance that meets standards on culpability-precipitation.".

itself culpable, my failure of foresight will be culpable, and so, accordingly, will my resulting irrationality.

It is not obvious that the notion of transparency can in fact be construed this way without losing the intuitive force behind internalism. It is certainly possible to think of the mind as containing, as it were, dark nooks and crannies capable of hiding things not apparent to a cursory glance by the inner eye but still accessible to careful scrutiny. The question is whether these metaphorical nooks and crannies can be *shallow* enough for their contents to remain within the subject's cognitive perspective, yet *deep* enough to leave room for the subject to fail to take inventory of their contents.

In general, if the picture is not to render the internalism unrecognizable, it must not represent access to the "hidden" material as contingent on the cooperation of the epistemic environment. That access must be entirely due to the subject. It may take some deliberate effort to reveal it, but if this effort is expended, the material is guaranteed to be revealed.

For an analogy, consider adding two large numbers by column addition. Provided that the calculation is carried out with sufficient rigor and care, the answer is more-or-less guaranteed to be forthcoming: there are very few outside factors that could prevent a sufficiently motivated investigator from, eventually, getting the right answer. Nevertheless, the investigation required is arduous and time-consuming, so someone may very well neglect it. Perhaps ascertaining the ir/rationality of one's beliefs is similar. Perhaps it requires a procedure that, while guaranteed to give the right answer, is arduous and time-consuming and which agents may therefore (culpably) choose to neglect. Or perhaps, for that matter, the procedure is not very arduous or time-consuming—perhaps it just involves the mental equivalent of casting your gaze in the right direction—but some agents nevertheless neglect it. Indeed, as remarked above, some authors, such as Rudy-Hiller (2017), deny that the benighting event has to be a deliberate *act*, like neglecting to perform a procedure. If these authors are right, it may suffice if the agent *forgets* to perform the procedure, as long as the forgetting meets relevant standard for culpability.

The question now becomes: Is there a plausible picture of how the mind works that can vindicate the above suggestion?

The picture cannot, I take it, be that upon forming a new belief, we (as it were) look through our existing beliefs one by one to check, for each one of them, whether they can be rationally conjoined with the candidate new belief. First off, it is unlikely that our beliefs are neatly individuated in the way required by this picture. Second, even if they were, a normal person has so many beliefs that it would be an unreasonable demand to make on people that they should look through all their existing beliefs before forming a new one. This latter point is reinforced by the consideration that irrationality can presumably reside not just in *pairs* of beliefs but in groups of three, four etc. beliefs. If the subject is required to compare a candidate new belief against all such *combinations* of beliefs, we have a combinatorial explosion on our hands. Third, and most importantly, there is, as far as I can tell, nothing in subjective experience that speaks in favor of the picture, no phenomenological evidence of this kind of deliberate, conscious act of scanning through one's belief set item

by item—and the act would have to be deliberate, rather than automatic and unconscious, to be something that could be neglected or forgotten.

A much more plausible picture is something like this: when we consider a candidate belief, our minds immediately serve up a number of other accommodations to our belief-set that we would have to make in order to rationally take the candidate belief on board, as determined automatically and subpersonally. But this immediate delivery of our cognitive subconscious might not be wholly reliable or exhaustive, and so, in order to ensure that we do not violate the requirements of rationality in taking on board the belief, we must let our thoughts wander across different subject-matters, more or less closely related to the issue at hand; try to draw out implications of existing beliefs as well as of the candidate new belief; and so on—and trust in the same subpersonal processes to alert us to any threatening conflicts.

The problem with *this* picture is that it once again entrusts the subject's epistemic access to the determinants of the ir/rationality of her beliefs to the cooperation of the epistemic environment. The proposed procedure can proceed indefinitely. There is no obvious point at which I can look back on my work and regard it as completed. Returning to our analogy: in the case of adding by column addition, there is a clear, definite point when I have gone through all the columns and can definitively ascertain that my work is done. The process of rumination described above, to the contrary, has no obvious endpoint. Thus, it again undermines transparency. I simply have no guarantee that, as long as I carry out the procedure diligently, I will eventually get the answer. The mind is, again, shrouded in opacity.

4.2 Higher-Order Ignorance

Recall: we are looking for ways for the internalist to accommodate the possibility of a subject 1) violating rationality *without* knowing that they're doing it while still 2) being culpable for the violation. Above, I discussed and rejected the suggestion that an appeal to *introspective effort* could do the job. In this section, I will examine the prospects for accommodating this possibility by appealing to *higher-order ignorance.*

Internalism, recall, is the view that the ir/rationality of a subject's beliefs depends only on facts that are transparent to her. For instance, if rationality prohibits contradictory beliefs, so that the ir/rationality of a subject's beliefs depend in part on whether they are contradictory, transparentism entails that it is transparent to a subject whether they are contradictory. Moreover, provided that my above argument against the appeal to introspective effort is sound, a subject who formed a contradictory set of beliefs couldn't be culpably ignorant of their contradictoriness.

But perhaps, even if a subject *knows* that her beliefs are contradictory, she can be culpably ignorant of the fact that that *this makes them irrational.* Perhaps she can be culpably ignorant of the fact that *irrationality prohibits contradictory beliefs*. Generalizing, it might be that a subject can be aware of those facts about their beliefs that *in fact* render them irrational, while remaining culpably ignorant of *what rationality requires*. In other words, perhaps a subject can be culpably ignorant of her own irrationality due to culpable *higher-order ignorance* of the requirements of rationality.

Consider, for instance, a dialetheist like Graham Priest, who holds that there are true contradictions and, accordingly, that it is rationally permissible to believe some contradictions (Priest, 1998). Suppose that, *contra* Priest, it is in fact rationally impermissible to believe any contradictions. Suppose that a certain dialetheist forms a contradictory set of beliefs, fully aware that these beliefs are contradictory but culpably ignorant of the fact that contradictory beliefs are rationally prohibited. She might then be culpable for the resulting irrationality.

To be *culpably* ignorant of something, I suggested above, requires that one's ignorance is due to a "benighting event" that is itself culpable. The word "benighting" evokes something like neglect: a failure to make use of a readily available epistemic resource. This picture sits ill with the example of the dialetheist, however. Whatever else one can say about Graham Priest, one cannot criticize him for having neglected to investigate the requirements of rationality. He, if anyone, has done his epistemic due diligence.

If the dialetheist's purported ignorance of the requirements of rationality results from a culpable benighting event, that event must therefore be construed in some other way. Perhaps the dialetheist's mistake lies in having overintellectualized the matter, seeking answers in recondite logical theorizing and thereby making herself insensitive to that inarticulate gut feeling that is the appropriate guide to what rationality requires. Or so the internalist could argue.

Perhaps there is something to be said for this "gut feeling" theory of the epistemology of the rational requirements. Even so, the resulting picture of what culpable irrationality consists in must strike us as strange. We might have thought, naïvely, that careful ratiocination was the paradigm of rationality and that the price for heeding the gut's siren call was to founder on the cliffs of irrationality. On the picture currently under consideration, the opposite turns out to be the case: it's the gut that gets it right, while the ratiocinators risk falling into (culpable) irrationality. That may be a gratifying picture to some, but it is a picture in which rationality is barely recognizable.

Be that as it may: the strategy of appealing to higher-order ignorance is not wedded to the example of the dialetheist specifically. Perhaps the internalist could say that culpable higher-order ignorance results from some much more straightforward epistemic failure. She then owes us an account of this failure.

In providing that account, an important choice-point for the internalist will be whether the requirements of rationality *themselves* are transparent to the subject. Call the claim that they are "higher-order internalism."

Suppose higher-order internalism is false. Then it would seem like internalism is also false. Internalism, recall, is the view that subjects have transparent access to those facts on which their ir/rationality depends. But the ir/rationality of a subject's beliefs, it would seem, depends not only on features of those beliefs but also on what rationality requires. After all, had rationality (perhaps *per impossibile*) required something other than what it in fact requires, then a subject who is in fact irrational might have been rational or vice versa.

Perhaps the internalist can wriggle out of this line of argument. She could try to say that though a subject's irrationality does perhaps in *some* sense depend on the requirements of rationality, this is not the sense of "depend" intended in

the definition of internalism. Such a move would not be wholly ad hoc. There is something that rings true in the claim that the sense in which a subject's rationality "depends" on features of those beliefs differ from the sense in which it "depends" on what rationality requires.

The move would not be completely ad hoc, perhaps, but it would still result in a view bereft of the essential features of internalism. The internalist, as we've seen, is concerned that a subject should be able reliably to translate her ambition to be rational into actual success in that endeavor, without having to trust in the cooperation of the epistemic environment. On the view currently under consideration, that will not be the case. Try as she might, the subject may still fail to be rational because she has been misinformed about what rationality requires by inclement epistemic circumstances. Adopting this revision would thus, at best, be a pyrrhic victory for the internalist. It would concede to the externalist the crucial point that rationality is sometimes contingent on the cooperation of the epistemic environment.

Suppose, instead, that higher-order internalism is true. Now, this assumption may strike the reader as *prima facie* implausible. There seems to be very little to recommend the idea that we have transparent access to the requirements of rationality. Witness, as evidence to the contrary, the massive philosophical literature debating what those requirements are (cf. Goldman, 1999, 287).

But perhaps this countervailing data can be explained away. After all, many philosophical controversies lend themselves to the suspicion of being mere verbal disagreements, and the controversy in question is certainly no exception (cf. my remarks above about the slipperiness of the term "rationality"). Perhaps, then, philosophical disagreement over what rationality requires is no more than disagreement over the semantics of the term "requirement of rationality." Meanwhile, the internalist would not seem to be committed to the implausible view that subjects have transparent epistemic access to the semantics of this term *as such*. She would seem to be committed only to the view that subjects have transparent access to what rationality *in fact* requires of them, regardless of what verbal label, if any, they attach to those requirements.

Given that we have such possibly-inarticulate access to what rationality requires, is there room for the type of benighting event that issues in culpable ignorance? The internalist would have to say more about what our transparent access to the requirements of rationality looks like before we can say for sure. However, as a general point, the dialectic for the first-order case will tend to repeat itself here on the higher level. For there to be room for benighting events, transparency cannot mean luminosity or self-intimation; there has to be some steps involved to *get* the relevant access. At the same time, the steps involved must be wholly up to the subject herself, not contingent on external factors. Thus, the question becomes, again, whether the resulting epistemological picture is plausible.

Here, I think it's clear that there's no phenomenological evidence of the requisite kind of conscious act, one that yields introspective but inarticulate access to what rationality requires and which could be neglected or forgotten.

Of course, even if subjects did have at their disposal the requisite kind of act, it would be difficult for us to recognize it *as such*. Since the access would be

inarticulate, we wouldn't necessarily know that what we were doing in performing the act was accessing the denotatum of the term "rational requirements." But presumably, the requirements would, when accessed, have to be recognizable *somehow*—if not as the denotatum of that term, then at least as a body of rules or strictures exerting normative pressure on belief-formation. Otherwise, it would be unclear in what sense the procedure resulted in knowledge *about what rationality requires.*

If the subject had such an act at their disposal, and the subject abstained from performing it for whatever reason, she might be culpably irrational. But I think it's clear that no such act is available. I, at least, wouldn't even know where to begin, whence to turn my mind, in order to gain my purported transparent access to the requirements of rationality. Rather, it seems to me that insofar as I am capable of conforming my beliefs to any normative strictures, my appreciation of those strictures is manifest directly in my appreciation of the first-order relations among the beliefs themselves. For instance, I believe Paris is north of Madrid. Were someone to claim that Paris is south of Madrid, I would appreciate that the belief I was thereby invited to form would contradict my already-held belief, and this appreciation *itself* would press me to decline the invitation. There is no separate act of becoming aware of the irrationality of contradictory beliefs, one I could neglect or forget to perform and so unwittingly add the new belief alongside my old one in full awareness of their contradictoriness.

5 Wanton Irrationality

We turn then to the second possibility for accommodating culpable irrationality canvassed at the end of Sect. 3: that the subject violates the requirements of rationality *wantonly*, i.e., in full knowledge that this is what she's doing—and is culpable for that act.

A culpable act of wanton irrationality would need to meet general requirements on culpable belief-formation. We established above that culpability for an act requires that the act be *due to* the subject. We also saw that this "due to" admits of different analyses, some of which (in terms of voluntary action) are less amenable to the possibility of culpable belief-formation than others.

I believe we should be able to agree that, at a minimum, a culpable act must be the product of a subject's capacity to respond to reasons. By this, I mean: she must be performing it due to her *having* or *taking* herself to either 1) have sufficient reason to do it or 2) lack sufficient reason not to do it (the latter might characterize acts performed on a mere whim); and perform it *for* those (would-be) reasons.[10]

[10] In proposing this criterion, I'm not taking myself to be saying anything controversial. The idea that responsibility requires responsiveness to reasons is associated with, and has been developed in most detail by, compatibilists (e.g. Fischer 1994; Fischer and Ravizza 1999; McKenna 2013; Sartorio 2016, chap. 4), but reasons-responsiveness should be acknowledged as a *necessary* condition for responsibility by everyone (Clarke 2003, 15 ff.).

Prima facie, it's difficult to see how a wanton violation of rationality *could* satisfy this requirement. To *wantonly* violate rationality, again, is to violate rationality in full knowledge that this is what one is doing. But it's hard to see how one could know fully that one's act will violate rationality and *not*, by that very fact, take oneself to have sufficient reason *not* to do it. One attractive view about the relationship between reasons and rationality is that to act on the reasons one takes oneself to have is to act "sub specie rationalitatis," i.e., according to what one takes to be rational. If this is true, then someone who *knew* that a certain course of action would be irrational yet acted thus anyway would *ipso facto* not be acting on the reasons she took herself to have and would thus not be satisfying the above requirement on culpable action.

With regard to acts of belief-formation, our present concern, this line of thinking is most plausible if restricted to *epistemic* reasons. If I *knew* that some act of belief-formation of mine would violate rationality, then I would presumably thereby also take myself to have sufficient epistemic reason not to perform the act. Any epistemic reason, real or perceived, in favor of the act would presumably be trumped by my knowledge that the resulting beliefs would be irrational. Denying this amounts to asserting that we can have sufficient epistemic reasons to form irrational beliefs, and that would seem to make a mockery out of the notions of epistemic reason and rational belief. It is less clear, perhaps, that a subject couldn't take herself to have sufficient reason to wantonly violate rationality if we are allowed to also count *practical* (including moral) reasons. It is controversial whether we *are* allowed to count practical reasons, i.e., whether one can have practical reasons for or against forming beliefs.[11] Even if we are, that does not suffice to show that we can have practical reasons to wantonly violate rationality. The typical claim by defenders of practical reasons for beliefs is not that we sometimes have practical reasons to violate rationality, but that pragmatic or moral considerations sometimes bear on the rationality of beliefs.

And even if we sometimes had practical reasons to violate rationality, that does not yet show that we can *take ourselves* to have practical reasons to *knowingly* violate rationality. For instance, Sanford Goldberg (2022) argues that we are sometimes morally obligated to believe contrary to our evidence, when that evidence is "contaminated" by structural injustice. The situations Goldberg describes, if they exist, would arguably constitute tradeoffs between morality and doxastic rationality. However, they still wouldn't constitute situation where we *take ourselves* to face a tradeoff between morality and rationality. If a subject is *aware* that her evidence is contaminated by structural injustice, that would presumably diminish its value *as evidence* and hence the rationality of believing in accordance with it. Only if the subject is *unaware* of the contaminated status of her evidence, can there truly be a conflict between morality and rationality. But if the subject is thus unaware of those moral reasons against following her evidence, she presumably cannot act *for* those reasons.

The internalist could offer outlandish scenarios of the following kind: an evil demon with mind-reading powers tells you to violate the requirements of rationality

[11] Recent interventions in this debate include McCormick (2014) and Reisner (2018).

(e.g., believe a contradiction), or he will kill you. In this scenario you will arguably take yourself to have sufficient reason to knowingly violate rationality (Reisner, 2011; Wedgwood, 2017, 33–35).

A number of things can be said about this idea. First, it is far from clear that this type of scenario is (conceptually or metaphysically) possible. As already noted, it seems a plausible principle that, insofar as you take yourself to have sufficient reason to do something, that is also, *ipso facto*, what you take to be rational. If this principle is valid, there could be no instruction the demon could give you that would be both 1) one you took yourself to have sufficient reason to follow and 2) one you knew would constitute a violation of rationality. Even if the principle is rejected, the outlandishness of the scenario speaks against at least its nomological possibility (though it might be possible, of course, to conceive of less outlandish scenarios).[12]

Second, even if such scenarios are *possible*, it must also be shown that you would actually be *culpable* if you acted on the demon's wishes, and it may be thought that the extreme stakes of the scenario are precisely the kind of thing that could exculpate a violation of rationality.

Third: To serve the internalist's needs, the scenario would have to be such that the subject not only takes herself to have sufficient reason to violate rationality but also *is capable* of actually violating rationality *for* those reasons. Recall that in the present context, a violation of rationality will be an act of belief-formation (or -revision or -retention). The internalist thus confronts the following question: is it at all possible to form *any* beliefs, irrational or not, for practical reasons?

To answer this question, we must attend to a conventional distinction between two ways in which someone might be loosely described as forming a belief for practical reasons: 1) forming a belief *directly* in response to practical reasons; and 2) taking action, for practical reasons, that has the formation of a certain belief as a *consequence*.

Begin with (1). The possibility of forming a belief directly in response to practical reasons is, I take it, precisely the kind of thing people deny when they deny the doctrine of doxastic voluntarism. To act for practical reasons is the domain of the will, and beliefs cannot be willed. If the internalist claims otherwise, she faces a dialectical uphill battle.

As to (2): Most philosophers, me included, would agree that this kind of *indirect* belief-formation in response to practical reasons is possible. For instance, I may have a practical reason to believe that my partner is faithful, despite strong indications to the contrary: believing this would ease my mind and improve my mood. Hence, I systematically avoid situations where I might confront evidence of my partner's unfaithfulness and seek out situations where I encounter evidence that she's faithful. As a consequence, I come to believe that she is faithful after all.

[12] Reisner (2011) also proposes a scenario where an eccentric millionaire offers you a huge cash reward for violating rationality. This is less outlandish than the demon scenario: it is at least within the realm of the clearly nomologically possible. But I would claim that this scenario is not actually one where you have sufficient practical reason to be irrational. After all, eccentric millionaires can't read minds, so in this scenario, you're better off *pretending* that you believe a contradiction but not actually doing it. And if we equip our millionaire with mind-reading powers, we're back in outlandish territory.

Examples of this kind hinge crucially, for their plausibility, on the detail that when I actually come to the point of forming my belief (e.g., that my partner is faithful), that belief-formation act itself occurs in a completely normal way as a response to my *epistemic* reasons. The practical reasons have played their entire role earlier in the process, in influencing the behavior that causes me to acquire some epistemic reasons rather than others. Thus, this type of scenario cannot provide a true example of a subject violating rationality for practical reasons.

In conclusion: a large number of considerations speak against wanton irrationality. Taking them all together amounts, in my appraisal, to a strong argument against its possibility.

6 Conclusion

If my reasoning above is sound, internalism cannot accommodate the possibility of culpable violations of rationality. But culpabilism entails that at least some violations of rationality are culpable. So culpabilism entails that internalism is false.

I have admitted to and sought to accommodate the possibility that the word "rationality" names multiple phenomena. Strictly speaking, then, what the argument shows is that insofar as culpabilism is true of *some* of these phenomena, internalism cannot be true of them as well.

The view I have defended is negative: culpabilism is inconsistent with internalism. It is natural to ask the further question whether *externalism* (i.e., the negation of internalism) is compatible with culpabilism, or whether culpabilism is, rather, unsustainable *tout court*. My suspicion here is that externalism is likely to be able to accommodating culpable irrationality by accommodating culpable ignorance; for instance, in the form of neglect of epistemic duties (as proposed by Paulo Faria (2009); cf. the discussion of Pierre in Sect. 2). Whether this suspicion is ultimately sustainable cannot, however, be determined here.

Acknowledgements The idea for this paper was originally sparked in conversation with Nils Franzén. He, Anna Nyman, Andrew Reisner, and Olle Risberg graciously read and gave insightful comments on earlier drafts. In addition to these four, I have been greatly aided by conversations with Carl Montan and Maria Svedberg. Gratitude is also due to the participants at the Uppsala higher seminar in theoretical philosophy, the Umeå higher seminar in philosophy, the 2022 Swedish Congress of Philosophy (*Filosofidagarna*), the 2023 Uppsala-Zürich workshop on epistemic normativity, and two anonymous reviewers for Erkenntnis.

Funding Open access funding provided by Uppsala University. This research was supported by the Swedish Research Council (Award number 2021-06690) and Åke Wibergs stiftelse (Award number H21-0050).

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

References

Alston, W. P. (1986). Internalism and externalism in epistemology. *Philosophical Topics, 14*(1), 179–221.
Boghossian, P. (1994). The transparency of mental content. *Philosophical Perspectives, 8*, 33–50. https://doi.org/10.2307/2214162
Boghossian, P. (2011). The transparency of mental content revisited. *Philosophical Studies, 155*(3), 457–465. https://doi.org/10.1007/s11098-010-9611-3
Bonjour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy, 5*, 53–73. https://doi.org/10.1111/j.1475-4975.1980.tb00396.x
Boult, C. (2023). The significance of epistemic blame. *Erkenntnis, 88*(2), 807–828. https://doi.org/10.1007/s10670-021-00382-0
Broome, J. (2007). Is rationality normative? *Disputatio, 2*(23), 161–178. https://doi.org/10.2478/disp-2007-0008
Brown, J. (2004). *Anti-individualism and knowledge*. MIT Press.
Clarke, R. (2003). *Libertarian accounts of free will* (1st ed.). Oxford University Press.
Clarke, R. (2014). Negligent action and unwitting omission. In R. Clarke (Ed.), *Omissions: Agency, metaphysics, and responsibility.* Oxford University Press.
Clarke, Randolph. (2017). Blameworthiness and unwitting omissions. In D. K. Nelkin & S. C. Rickless (Eds.), *The ethics and law of omissions.* Oxford University Press.
Faria, P. (2009). Unsafe reasoning: A survey. *DoisPontos, 6*(2), 185–201.
Feldman, R., & Conee, E. (2001). Internalism defended. *American Philosophical Quarterly, 38*(1), 1–18.
Fischer, J. M. (1994). *The metaphysics of free will: An essay on control*. Blackwell.
Fischer, J. M., & Ravizza, M. (1999). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.
Ginet, C. (1975). *Knowledge, Perception, and Memory*. D. Reidel.
Glüer, K., & Wikforss, Å. (2009). Against content normativity. *Mind, 118*(469), 31–70. https://doi.org/10.1093/mind/fzn154
Glüer, K., & Wikforss, Å. (2013). Against belief normativity. In Timothy Chan (Ed.), *The aim of belief.* Oxford University Press.
Goldberg, S. C. (2022). What is a speaker owed? *Philosophy & Public Affairs, 50*(3), 375–407. https://doi.org/10.1111/papa.12219
Goldman, A. I. (1999). Internalism exposed. *The Journal of Philosophy, 96*(6), 271–293.
Kiesewetter, B. (2017). *The normativity of rationality*. Oxford University Press.
Kolodny, N. (2005). Why be rational? *Mind, 114*(455), 509–563. https://doi.org/10.1093/mind/fzi509
Kripke, S. (1979). A Puzzle About Belief. In A. Margalit (Eds.) *Meaning and Use: Papers Presented at the Second Jerusalem Philosophical Encounter,* (pp 239–283). D. Reidel.
Lord, E. (2018). *The importance of being rational*. Oxford University Press.
Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.
McCormick, M. S. (2014). *Believing against the Evidence: Agency and the ethics of belief*. Routledge.
McKenna, M. (2013). Reasons-responsiveness, agents and mechanisms. In David Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 1, pp. 151–83). Oxford University Press.
Millikan, R. G. (1993). White queen psychology; or, the last myth of the given. *White queen psychology and other essays for Alice* (pp. 279–363). MIT Press.
Millikan, R. G. (2001). Cutting philosophy of language down to size. *Royal Institute of Philosophy Supplements, 48*, 125–140. https://doi.org/10.1017/S1358246100010742
Nagel, T. (1979). Moral luck. *Mortal questions* (pp. 24–38). Cambridge University Press.
Owens, D. J. (2000). *Reason without freedom: The problem of epistemic normativity*. Routledge.

Priest, G. (1998). What is so bad about contradictions? *The Journal of Philosophy, 95*(8), 410. https://doi.org/10.2307/2564636
Reisner, A. (2011). "Is there reason to be theoretically rational?" In A. Reisner (Ed.), *In reasons for belief*. Cambridge University Press.
Reisner, A. (2018). Pragmatic reasons for belief. In D. Star (Ed.), *The Oxford handbook of reasons and normativity*. Oxford University Press.
Rudy-Hiller, F. (2017). A capacitarian account of culpable ignorance. *Pacific Philosophical Quarterly, 98*(1), 398–426. https://doi.org/10.1111/papq.12190
Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies, 114*(1/2), 47–79.
Sartorio, C. (2016). *Causation and free will* (1st ed.). Oxford University Press.
Smith, H. (1983). Culpable ignorance. *The Philosophical Review, 92*(4), 543–571. https://doi.org/10.2307/2184880
Sorensen, R. A. (1998). Logical luck. *The Philosophical Quarterly, 48*(192), 319–334.
Srinivasan, A. (2015). Normativity without cartesian privilege. *Philosophical Issues, 25*(1), 273–299. https://doi.org/10.1111/phis.12059
Steup, M. (2012). Belief control and intentionality. *Synthese, 188*(2), 145–163. https://doi.org/10.1007/s11229-011-9919-3
Stich, S. (1993). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. MIT Press.
Wedgwood, R. (2017). *The value of rationality*. Oxford University Press.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.