Rationality Reunified

Keshav Singh

Forthcoming in *Oxford Studies in Metaethics, vol. 20*

Introduction[1]

What is rationality? Increasingly, it seems difficult to provide a single answer to this question. This is because philosophers now distinguish between two kinds of rationality: substantive and structural. Roughly, substantive rationality consists in holding attitudes that are substantively reasonable or justified, whereas structural rationality consists in holding attitudes that fit together in the right ways.[2] Recognizing this distinction raises questions about the relationship between the two kinds of rationality.

On some views, the existence of two distinct kinds of rationality is merely apparent. Really, there is only a single distinct kind of rationality, and the other is either reducible to or eliminable in favor of the first. However, an increasingly popular view is *dualism*, which rejects attempts at reduction and elimination. Dualism is made attractive by the apparent difficulties faced by various monist views. But if dualism is true, and there really are two irreducibly distinct kinds of rationality, the worry is that there's no single answer to the question, "what is rationality?" The appropriate response to such a question would be, "It depends – which kind of rationality do you mean?"

There is something unsatisfying about this. Intuitively, the question "what is rationality?" *is* one to which there is some kind of single answer. There must be some

---

[1] Thanks to Lindsay Brainard, David Faraci, Daniel Fogal, David Horst, Mikayla Kelley, Alex Worsnip, an anonymous reviewer, and audiences at the 2020 Northeast Normativity Workshop, the 2023 St. Louis Annual Conference on Reasons and Rationality, the 20th Annual Madison Metaethics Workshop, and the 2024 APA Central Division meeting for helpful feedback on various iterations of this paper.

[2] The terminology of structural vs. substantive rationality has its roots in Scanlon, who distinguishes "*structural* claims about rationality" from "*substantive* claims about what is a reason for what" (2003, 84). I am uncertain where the exact terminology of structural vs. substantive rationality first appeared, but it is now standardly deployed in the theory of rationality.

characterization of rationality in virtue of which it makes sense to categorize both substantive and structural rationality as kinds of rationality in the first place. In other words, intuitively, there must be some underlying unity of rationality, even if substantive and structural rationality are in some sense distinct.

My goal in this paper is to explore how rationality might ultimately be unified even in absence of a reduction (or elimination) of one kind of rationality to the other. The account I will arrive at appeals to a certain picture of the nature of *rationally evaluable attitudes* like beliefs, fears, intentions, and preferences (hereafter just 'attitudes'), on which each attitude is a piece of our take on the world. Each attitude we hold involves a commitment to something about the world being a certain way. As such, our attitudes aim at getting things correct.

The basic idea behind my approach to understanding rationality is as follows. In virtue of the picture of attitudes described above, there is a fundamental normative standard of correctness for each attitude. Given our epistemic limitations, we cannot directly ensure satisfaction of this standard of correctness. But we can manifest our commitment to satisfying this standard by following the norms of rationality. This is how the norms of rationality are derived from the fundamental standard of correctness. Norms of substantive and structural rationality, on my view, correspond to two different ways of manifesting the commitment to getting things correct, given our inability to directly ensure that we do.

Here is the plan for the paper. §1 will consist primarily of ground clearing: I'll review the central motivations for preferring dualism to monism, as well as my account of the nature of attitudes that yields the fundamental norm of correctness. Then, I'll sketch accounts of substantive rationality (§2) and structural rationality (§3) that make

clear how the norms of each are derivative of the fundamental standard of correctness. In doing so, I'll show how rationality can be reunified (§4). In §5, I'll conclude with brief remarks about how being more committal about the nature of attitudes might shed further light on the normativity of rationality.

1. Ground-Clearing

1.1. Monism vs. Dualism

Until very recently, monism has been the dominant view in the growing literature on the theory of rationality. According to monism, there is ultimately only one kind of rationality, and any putative second kind of rationality is either eliminable or reducible to the first.[3] The most popular version of monism argues that structural rationality can be entirely explained (or explained *away*, for the eliminativist) in terms of substantive rationality.[4] This version of monism relies on what Worsnip (2021) and Fogal and Worsnip (2021) call the *Guarantee Hypothesis*. According to the Guarantee Hypothesis, any apparently structurally irrational set of attitudes is guaranteed to contain at least one attitude that is substantively irrational. Thus, the apparently distinct failure of structural rationality just comes down to a guaranteed failure of substantive rationality.[5]

---

[3] See Worsnip (2021, Ch. 3-4) for a helpful discussion of the differences between what he calls eliminativist and reductivist forms of monism.

[4] There are other versions of monism. For example, Broome (2007, 2020) argues for a version of monism on which only structural rationality is really rationality, and what has been called substantive rationality is simply not a kind of rationality at all. Though Broome's arguments are interesting, they have not garnered widespread support, and I don't have space to discuss them further here. For responses to Broome's view, see the commentaries on his 2020 paper, especially Arpaly (2020), Kiesewetter (2020), and Singh (2020). See also Worsnip (2010, Ch. 4).

[5] Of course, the reductivist version of monism entails neither that the property of structural (ir)rationality does not exist, nor that it is identical to the property of substantive (ir)rationality. Nevertheless, I think it remains fair to say that the rational failure at issue in structural irrationality is no longer distinct from that of substantive irrationality on a reductivist view. Thanks to Alex Worsnip for pressing me to clarify this.

This version of monism has been defended by prominent theorists of rationality, such as Kiesewetter (2017) and Lord (2018).[6] For these monists, rationality ultimately comes down to having attitudes that are substantively reasonable or justified. As with any monist view, this could take either a reductionist or an eliminativist form, and it isn't always clear which form monists mean to defend. But that won't matter for my purposes, as both forms are subject to many of the same objections.

Though the version of monism described above (hereafter just monism) remains prominent and popular, it has recently been subject to significant criticism from Worsnip (2021) and Fogal and Worsnip (2021). They defend dualism partly on the grounds that the Guarantee Hypothesis is false: not all cases of structural irrationality can be explained in terms of a guaranteed failure of substantive rationality. For example, take the case of cyclical preferences: preferring each of A to B, B to C, and C to A. This set of preferences seems clearly to be irrational in the way thought to be distinctive of structural irrationality. But arguably, it need not involve any substantive rational failure, because each individual preference could be substantively rationally permissible.

I won't say much more about the objections to monism here, since I think Fogal and Worsnip have already presented those objections quite persuasively. But I will say that I don't think it should be so surprising that attempts to establish that every case of apparent structural irrationality is ultimately a case of substantive irrationality run into trouble. Intuitively, the two kinds of rational failure can not only come apart, but also compound each other. Fogal and Worsnip argue persuasively for the existence of cases in

[6] Kolodny (2005, 2007, 2008) defends a precursor to this sort of view, but does not adopt the terminology of substantive vs. structural rationality.

which adding structural irrationality to a case of substantive irrationality, or vice versa, generates a strong intuition of additional irrationality. Moreover, even defenders of monism like Kiesewetter and Lord are forced to respond to some counterexamples by declaring that when the apparent structural irrationality can't be explained by substantive irrationality, the irrationality is only apparent. All of this makes a strong case against monism and in favor of dualism.

If we reject monism and accept dualism, though, it raises new questions. If there are really two distinct kinds of rationality, neither of which is reducible to the other, are they related in any deep way? Is there some meaningful sense in which the two kinds of rationality are still unified? On one version of dualism, the answer to these questions is *no*: there's nothing that ultimately unifies substantive and structural rationality. Perhaps the term 'rationality' is polysemous between the two, and the relation between the two meanings is some relatively shallow fact about both involving normative standards one is criticizable for failing to satisfy. Call this *strong dualism*. Worsnip seems to be sympathetic to strong dualism, writing in his initial introduction of dualism that "'Rationality' is a word with more than one meaning" (2021, 4).

On another version of dualism, the answer to the above questions is *yes*: there *is* something that ultimately unifies substantive and structural rationality. While the two kinds of rationality are importantly distinct, there is at bottom some deep commonality between them in virtue of which they count as members of the kind *rationality*. On such a view, the relationship between substantive and structural rationality would be stronger than mere polysemy. Call this *weak dualism*. The advantage of weak dualism over strong dualism is that it doesn't leave us with a fundamentally disunified picture of rationality, or without an answer to questions about what rationality really comes down to.

I'll defend a version of weak dualism in this paper. As I'll argue, rationality in general is the part of the normative realm that has to do manifesting the commitment to getting things correct, given our epistemic limitations. Substantive and structural rationality are each distinct kinds of rationality because they consist in norms that accomplish this in different ways. Each of them is a way of aiming to get things correct given our epistemic limitations, and neither of these ways of so aiming is reducible to the other.

1.2. Correctness

Before I can defend this unified picture of rationality, I must at least sketch the overall picture of the nature and normativity of attitudes against which I'll develop my account. In Singh (2022), I develop an account of the nature of attitudes on which an attitude is a special kind of intentional mental state that is constituted by a commitment on the part of its holder to the attitude's object having a certain property. For now, I will only briefly outline the notion of commitment I have in mind, though I will return to it and the end of the paper. For the purposes of developing my account of rationality, the important thing is that the fundamental norm of correctness for each attitude is *generated* by its constitutive commitments. In committing yourself to the attitude's object having a certain property, you make it the case that the object's actually having that property is the normative standard of correctness for that attitude.

Take the example of belief. Belief takes a proposition as its object and represents that proposition as true. But of course, not every representation of a proposition as true counts as a belief, as opposed to, say, a supposition. On my account of the nature of attitudes, what makes an attitude a belief is that it is constituted by a commitment on the

part of the believer to the truth of the proposition.[7] This commitment generates a fundamental norm of correctness for belief that is a norm of truth. Such a norm might be formulated as a prescription (believe *p* if and only if *p* is true) or in some other way. But however it's formulated, it will be a norm of correctness telling us to believe what is true.

Whether or not you accept my story about what gives rise to this norm, it's a commonly accepted view that the fundamental norm of belief is a truth norm, because what it is for a belief to be correct is for it to be true. If this is so, it seems natural to think that for each attitude, there is a fundamental norm of correctness that gets filled in with whatever the correctness condition for that attitude is. This would be neatly explained by my account of the nature of attitudes as commitments to their objects having certain properties, but one could accept it without accepting my account. Of course, one would then need some other sort of argument for the conclusion that each attitude has a correctness condition, which is something I cannot undertake in this paper.

The foregoing yields a picture on which the fundamental normative standard for attitudes is the following, which I will render as a prescription for simplicity's sake:

> **Correctness:** for any rationally evaluable attitude A with correctness condition C, hold A only if A satisfies C.[8]

I think it's quite plausible that all rationally evaluable attitudes have this standard of correctness as their fundamental norm (and, as I will argue shortly, that this explains why they are rationally evaluable). This is because I think what it is for some mental state to be an *attitude* towards something is for it to be a piece of your take on the world. The

---

[7] See Singh (forthcoming) for more on my account of belief in particular.

[8] I render this as an "only if" rather than a biconditional because I want to remain neutral on whether the standards of correctness for attitudes are always requiring norms or are sometimes merely permitting or recommending norms. Thanks to a reviewer for suggesting I clarify this.

easiest example is belief. When you believe *p*, a piece of your take on the world is that *p* is true. But as I will explore later, I think other attitudes have standards of correctness that can be drawn out from consideration of how they are pieces of our take on the world.

The problem with standards of correctness, however, is that we can't directly ensure satisfaction of these standards, given our epistemic limitations. You can't ensure that you attribute some feature to the world if and only if it actually has that feature, because you aren't omniscient. Because the fundamental norm of correctness isn't a norm we can directly follow, derivative norms fall out of it: norms with which we *can* directly comply, because they're constrained by our epistemic situation. These are the norms of rationality. Rationality is the epistemically constrained shadow of correctness.

This conception of rationality fits with what many epistemologists have already written about belief. For example, here is Wedgwood:

> …it is essential to beliefs that they are causally regulated by certain standards of rational or justified belief, and…the ultimate purpose or point of conforming to these standards is not just to have rational or justified beliefs purely for their own sake, but to ensure that one believes the proposition in question if and only if that proposition is true. (2007, 154)

In this passage, Wedgwood seems to endorse the idea that norms of rational or justified belief are derivative of a fundamental truth norm of belief.[9] And he characterizes this fundamental truth norm of belief as a norm of correctness. My hope is that this way of looking at things can be extended to all (rationally evaluable) attitudes: the norms of rationality for those attitudes are derivative of the fundamental norms of correctness for

[9] Relatedly, many epistemologists argue that norms of rational credence are derivative from accuracy norms. See, e.g., Schoenfield (2015).

those attitudes. This gives us a unified picture of what rationality is ultimately about: manifesting the commitment to getting things correct given our epistemic limitations.

Importantly, one could have different views about the derivation relation between epistemically unconstrained standards of correctness and the epistemically constrained norms of rationality. For example, one could understand the derivation relation as an instrumental relation, wherein the norms of rationality *promote* satisfaction of the standard of correctness. But such an understanding would have highly revisionary implications for rationality, especially in cases where rationality and correctness come apart. It may even entail some kind of consequentialist theory of rationality. For those interested in such a theory, instrumental derivations could be substituted into my larger account of rationality. But I won't understand the derivation relation in instrumentalist terms. Instead, I will understand it in terms of manifesting the commitment to getting things correct.[10] I will say more about how I understand the significance of commitment in §5.

2. Substantive Rationality

Substantive rationality is almost ubiquitously understood in terms of reasons. As it's usually put, substantive rationality consists in being properly responsive to your reasons, and substantive irrationality consists in a failure to properly respond to your reasons. Monists like Kiesewetter and Lord and dualists like Fogal and Worsnip largely agree on this gloss of substantive rationality. Although I basically agree with it too, a brief

[10] There are similarities between what I say here and Sylvan's (2018, 2020) non-instrumentalist understanding of derivative epistemic value in veritist epistemology. Thanks to David Horst for pointing this out, as well as for helpful suggestions on how to clarify my understanding of the derivation relation between norms of rationality and standards of correctness.

clarification is in order. The gloss of substantive rationality as *responsiveness* to reasons is implicitly a gloss of *ex post* rationality. Roughly, *ex ante* rationality is a matter of what attitudes would be rational for you to hold, whereas *ex post* rationality is a matter of whether you hold attitudes *rationally*. This corresponds roughly to the distinction between being subject to a rule and *following* that rule.

If *ex post* rationality is a matter of responsiveness to your reasons for an attitude (which corresponds to following a rule) *ex ante* rationality must be a matter of that attitude's simply being supported by those reasons (which corresponds to simply being subject to a rule, whether or not you follow it). In trying to explain what substantive rationality is, we're trying to explain a norm that applies to our attitudes in virtue of which failure to comply with that norm would subject us to the criticism that we are (substantively) irrational. So, we should first and foremost explain what *ex ante* rationality is. And *ex ante* substantive rationality should be understood as having to do with which attitudes are *supported* by your reasons, rather than whether you respond to those reasons.

Of course, not just any reasons are relevant here. Philosophers standardly distinguish between objective and subjective reasons. It's clear that if substantive rationality is a matter of getting things correct given our epistemic limitations, the relevant kind of reason can't be objective reasons.[11] Moreover, as Fogal and Worsnip (2021) have persuasively argued, the category of 'subjective reasons' is too coarse-grained. In place of the objective/subjective distinction, they propose a tripartite

[11] Or at least, it can't be *all* objective reasons. Lord (2018) argues that the reasons relevant to substantive rationality are a subset of objective reasons that he calls 'possessed reasons.'

distinction between fact-relative, evidence-relative, and belief-relative reasons. Fact-relative reasons are maximally objective, and belief-relative reasons are maximally subjective. But evidence-relative reasons are somewhere in between: they are considerations that count in favor of an attitude given your evidence (as opposed to given your beliefs or given the facts).

As with much of what Fogal and Worsnip argue, I won't rehash their arguments here. I'll simply take on their tripartite distinction, as I find it completely persuasive. Based on this tripartite distinction, Fogal and Worsnip argue that the reasons relevant to substantive rationality must be evidence-relative reasons. They argue this based on the fact that one's belief-relative reasons could support responses that are clearly intuitively irrational, such as intending to drink petrol you believe is gin despite all your evidence indicating it's petrol. I find this completely persuasive as well. Out of the three kinds of reasons, the process of elimination shows that evidence-relative reasons must be the reasons relevant to rationality.

As Worsnip puts it in his book, "substantive rationality consists in correctly responding to evidence-relative reasons" (2021, 46). Given earlier discussion, this is best understood as a description of *ex post* substantive rationality. For some attitude to be *ex ante* substantively rational, then, would be for it to be supported by one's evidence-relative reasons. With this on the table, we can formulate more precise norms of substantive rationality for both rational permission and requirement:

> **Substantive Rational Permission (SRP):** You are (substantively) rationally permitted to hold attitude A if and only if you have sufficient evidence-relative reasons to hold A.
>
> **Substantive Rational Requirement (SRR):** You are (substantively) rationally required to hold attitude A if and only if you have decisive evidence-relative reasons to hold A.

These norms of substantive rationality give us the tools to explain which attitudes are *ex ante* rational in terms of your evidence-relative reasons. Of course, explaining whether the attitudes you hold are *ex post* rational would require something more: an account of what it is to correctly respond to these reasons. And that is a subject for another paper.

The next step is to explain how these norms of substantive rationality are derivative of the fundamental correctness norm. To do so requires another piece of my overall picture of normativity, which is that reasons of the right kind for attitudes are considerations that bear on whether the attitude is correct. This is what makes them right-kind reasons, as opposed to wrong-kind reasons, which bear on the value of holding the attitude.[12] This is an independently plausible way of drawing the distinction between right-and wrong-kind reasons. And it's important here because substantive rationality depends only on right-kind reasons.[13]

To return to the example of belief: what makes a consideration that counts in favoring of believing *p* a right-kind reason for that belief? Plausibly, it's that that consideration bears in some way on whether or not *p* is true. This is certainly true of evidential reasons, which are paradigmatic right-kind reasons for belief. By contrast, incentives to believe *p*, which are paradigmatic wrong-kind reasons, bear not on whether *p* is true, but on the value of believing *p*. This can be explained by the view that right-kind reasons for attitudes are considerations that bear on whether the attitude is correct.

---

[12] For examples of correctness-based accounts of the distinction between right-kind and wrong-kind reasons for attitudes, see Danielsson and Olson (2007), Schroeder (2010) and Sharadin (2016).
[13] For an argument, see Worsnip (2021, 44-46).

Now, let's return to the idea that norms of rationality are about manifesting the commitment to getting things correct, given our epistemic limitations. Given our epistemic limitations, how can we attempt to comply with a norm prescribing correct attitudes if not by being guided by the considerations that, according to our evidence, bear on whether those attitudes are correct? Right-kind evidence-relative reasons just are such considerations. Therefore, if derivative, epistemically constrained norms fall out of our inability to directly follow the correctness norm given our epistemic limitations, the norms of substantive rationality already identified are exactly what we should expect to end up with. The norms we have independently identified as the norms of substantive rationality turn out to be explained by the fundamental correctness norm in exactly the way I've suggested they would.

The above sketch of substantive rationality makes up half of the unified picture of rationality I aim to defend. I've claimed that rationality is ultimately about manifesting the commitment to getting things correct given our epistemic limitations. If substantive rationality is a matter of what our evidence-relative reasons call for, and our evidence-relative reasons are considerations that point us toward correct attitudes given our evidence, then it's clear how the norms of substantive rationality manifest the commitment to getting to getting things correct given our epistemic limitations.

3. Structural Rationality

In this section, I'll argue that the norms of structural rationality are also derivative of the fundamental norm of correctness. This will be a somewhat more complex endeavor, as theorists of rationality have identified a large number of particular norms of structural rationality. These include norms of belief consistency, intention consistency, means-ends

coherence, enkrasia, and many others. It's a matter of significant debate what unifies even these more specific norms of structural rationality. So, my goal in this section will be to first draw out the fundamental norm of structural rationality, before explaining in §4 how it's derivative of the fundamental norm of correctness. As such, I'll briefly consider two recent accounts of structural rationality, from Worsnip (2021) and Brunero (2020), before sketching my own account.

3.1. Worsnip and Brunero

According to Worsnip, the norms of structural rationality take the form of prohibitions on incoherent sets of attitudes. To explain what unifies instances of incoherence, Worsnip presents a thesis he labels *Incoherence Test:*

> A set of attitudinal mental states is jointly incoherent iff it is (partially) constitutive of (at least some of) the states in the set that any agent who holds this set of states has a disposition, when conditions of full transparency are met, to revise at least one of the states. (2021, 133)

I agree with Worsnip that the norms of structural rationality are prohibitions on incoherent sets of attitudes. And I find Incoherence Test plausible as a test for incoherence. But I find it less plausible that it's anything more than a test. Indeed, Worsnip himself commits to Incoherence Test only as a test and is "neutral on whether this property is identical to the property of incoherence" (128).[14] It seems to me that to uncover the ultimate nature of incoherence, we need to know what it is about the nature of attitudes in virtue of which we're disposed to revise them in cases of incoherence. The

[14] In earlier work, Worsnip (2018) presents Incoherence Test as a metaphysical thesis about incoherence rather than just a test.

dispositions Worsnip identifies don't seem like they can be at the bottom of the explanation.

Brunero's account focuses specifically on *instrumental* rationality and doesn't commit to an account of what unifies all norms of structural rationality. Brunero argues that instrumentally incoherent sets of attitudes are ones that involve guaranteed *constitutive aim failure*. As he understands constitutive aims, they are the "job descriptions" of attitudes within our psychology (177). For example, the constitutive aim of belief is truth because the job description of belief is to correctly represent how things are. The coherence of a set of attitudes is a matter of whether it's possible, given the logical relations among the contents of your attitudes, for all of the attitudes to achieve their constitutive aims.

Brunero's view is thus similar to the Guarantee Hypothesis discussed and rejected by Fogal and Worsnip, but in terms of constitutive aims rather than substantive rationality. Two further differences between Brunero's view and the Guarantee Hypothesis are important. First, Brunero denies that instrumental rationality is genuinely normative on his account. While we can speak of requirements of instrumental rationality, they aren't backed by reason or necessarily tied to criticizability. Second, Brunero denies that his account can unify all instances of incoherence. In precisely the kinds of cases that Fogal and Worsnip use to object to the Guarantee Hypothesis, such as intransitive preferences, Brunero concedes that his account is "of no help" (205). Because his account is neither a normative account nor a unified account of the entirety of structural rationality, it can't figure in an attempt to unify structural and substantive rationality.

Brunero's account is helpful, though. He argues that a set of inconsistent beliefs is incoherent *not* because it's guaranteed to contain a substantively irrational belief, but because it's guaranteed to contain a *false* belief. While Brunero describes this as a failure of belief to achieve its constitutive aim, we can also look at it as a failure of belief to satisfy its constitutive correctness condition. A plausible explanation of the norm of belief consistency is that it prohibits sets of inconsistent beliefs because such sets are guaranteed to contain *incorrect* beliefs. This suggests another version of the Guarantee Hypothesis that diverges from both the Kiesewetter-Lord approach and Brunero's by focusing on correctness.

3.2. Guaranteed Incorrectness: Easier Cases

Consider a version of the Guarantee Hypothesis that focuses on incorrectness instead of substantive irrationality. I think we can make progress toward unifying the more specific norms of structural rationality with something like the following:

> **Guaranteed Incorrectness (GI):** If some set of attitudes S is such that, in virtue of the relations between the constitutive correctness conditions of (at least some of) the attitudes in S, S is guaranteed to contain at least one incorrect attitude, then it is (structurally) rationally prohibited for you to hold S.[15]

I think GI clearly can explain many of the norms of structural rationality. The prohibition on inconsistent beliefs is the most obvious case: if a set of beliefs is inconsistent, it's impossible in virtue of the correctness condition for belief for all the beliefs in the set to be true.

---

[15] Aside from Brunero, similar approaches to GI are pursued by Fullhart and Martinez (forthcoming), and especially Fink (forthcoming).

Let's turn to some other cases. It's standardly thought that there is a prohibition on inconsistent intentions – intending to φ, intending to ψ, and believing one cannot both φ and ψ. How GI explains this depends on what the correctness condition for intention is. Here, it will be helpful to return to the account of the nature of attitudes I presented earlier. On that account, each attitude is a piece of your take on the world. So, we may ask, what piece of your take on the world does your intention to *φ* correspond to? What are you committing to about the world when you intend to *φ*?

While I can't give a complete account of intention here, it seems to me there are at least two things you're committing to about the world when you intend to *φ* (which thus constitute correctness conditions for intending to *φ*). The first is that *φ*ing is choiceworthy. The second is that you are actually going to *φ*.[16] Each of these commitments fits with independently plausible views about intention. The first fits with the Guise of the Good thesis but requires endorsing only a very weak version of it. The second fits with either cognitivism about intention, or non-cognitivist views on which intention aims at "making its content true" (Bratman 2018, 42).[17] It's the second commitment that does the work in explaining the prohibition on inconsistent intentions. If you can't both φ and ψ, then it can't both be the case that you're going to φ and that you're going to ψ. Thus, at least one of the three attitudes in the set is guaranteed to be incorrect, even if they are each individually choiceworthy.

---

[16] Perhaps it's even plausible to hold that what you're committing to is that you're going to *φ because φ*ing is choiceworthy. This would address a potential worry that I'm conceiving of intention as objectionably disunified in its commitments.

[17] Thanks to Mikayla Kelley for drawing my attention to the fact that Bratman-style non-cognitivism fits with my view as well as cognitivism.

GI will yield a similar explanation of means-ends coherence. Means-ends coherence prohibits the combination of intending to φ, believing you're only going to φ if you ψ, and intending not to ψ. For all of these attitudes to be correct, it would have to be the case that you're not going to ψ, you're only going to φ if you ψ, and yet you're going to φ. This is impossible. So, as long as part of what it takes for an intention to φ to be correct is that you're going to φ, GI can explain both the norm of intention consistency and the norm of means-ends coherence.

Given my assumptions about the correctness conditions for intentions, GI can also explain enkrasia. A norm of enkrasia prohibits believing that one ought to φ and intending not to φ. It also prohibits believing that one ought *not* to φ and intending to φ. GI can explain both of these by adverting to the other aspect of intention's correctness, choiceworthiness. This is because, plausibly, it can't be the case that you ought to φ and yet φing is not choiceworthy. Neither can it be the case that you ought not to φ and yet φing *is* choiceworthy. Ought and choiceworthiness are too closely related for either of those to be possible.

At this point, you might note that I've formulated all of these norms as prohibitions on sets of attitudes, whereas norms of structural rationality are often formulated as prohibitions on sets including both attitudes and *absences* of attitudes. I've done this because it's simpler to show how GI applies in cases of sets of attitudes. In cases where the apparently structurally irrational set includes the absence of an attitude, there are two different ways to go. The first would be to deny that such cases are necessarily structurally irrational, on the grounds that one's take on the world is merely incomplete, rather than positively incoherent. The second way of going would be posit some auxiliary principle according to which absences can ground the structural irrationality of a set just in case

any other committal option besides the attitude one lacks would create a set of attitudes that's structurally irrational according to GI. I'll remain neutral on which strategy is preferable.

3.3. Guaranteed Incorrectness: Harder Cases

So far, I've shown how GI provides a unified explanation of several norms of structural rationality. However, I haven't yet addressed the cases that lead dualists to reject the Guarantee Hypothesis. Since GI is a version of the Guarantee Hypothesis, we should expect such cases to provide a stress test for GI. I'll focus on two examples here: cyclical preferences, and what Worsnip calls inter-level coherence.

3.3.1. Cyclical Preferences

Importantly, Worsnip presents the case of cyclical preferences as a case of what he calls *permissive* counterexamples to the Kiesewetter-Lord version of the Guarantee Hypothesis (2021, 68-69). The idea is that for each of preferring A to B, preferring B to C, and preferring C to A, the reasons are sufficient to make each preference substantively rationally permissible, but the reasons are not decisive such that any of them is substantively rationally required. If this is possible, Kiesewetter and Lord can't explain why cyclical preferences are structurally irrational. Moreover, Worsnip argues that their view can't explain what he calls *permissive cases* more generally.

To see how GI fares here, we need to think more carefully about two things: the correctness conditions for preference, and what explains why the preferences are permissive in Worsnip's case. Let's start with the latter by looking at how Worsnip describes the case:

> Suppose that one is at a restaurant and that there are three available dishes – kale Caesar salad, tacos, and ma po tofu – each delicious and appealing in its own, different way. For any pair of these three dishes, each will have *some* good features that the other lacks. Let's also stipulate that you're aware of all these features, so that they constitute evidence-relative reasons for preference. (2021, 68-69)

What makes it the case that the reasons for each of these preferences merely *permit* preference, as opposed to requiring it? As reasons for preference, Worsnip cites considerations like "the salad is healthier than the tacos" and "the tacos are most robust and spicy in flavor than the salad." Why is it the case that these reasons don't simply add up to determine a uniquely rational preference ordering of salad, tacos, and tofu?

Here's a plausible hypothesis: these sorts of considerations don't have fixed, context-independent weights. Rather, their relative weights seem to depend on facts about you that can change depending on the context, such as what you're in the mood for or what you're prioritizing at the time. In any given pairwise comparison, one could prioritize certain things, making it substantively rational to prefer one way or the other. This explains why preference is permissive in such cases.

Now consider what makes preferring A to B correct. As with other attitudes, I can't offer a full theory of preference here, so let's operate with a toy theory according to which the correctness condition for preference is relative choiceworthiness.[18] In other words, it's correct for you to prefer A to B if and only if A is all-things-considered worth choosing over B. Next, consider the gloss on right-kind reasons I offered earlier: right-kind reasons are considerations that bear on whether the attitude is correct. If the explanation of the permissiveness of preferences is that the weight of your (right-kind) reasons for

[18] Whatever the actual correctness condition is, it will plausibly be some comparative property that is transitive and asymmetric.

preferences depends on further facts about your psychology at the time, then the correctness of these preferences must also depend in part on these facts. That is, the ranking of relative choiceworthiness between the three options must depend in part on the significance you assign to the various features of the three options that bear on whether they are all-things-considered worth choosing over other options.

If all of this is right, then it's impossible for A to be all-things-considered worth choosing over B, B to be all-things-considered worth choosing over C, and C to be all-things-considered worth choosing over A, *even in permissive cases*. Even if each preference is substantively rationally permissible, once the criteria for relative choiceworthiness in a particular deliberative episode are settled, it's impossible for every preference in the cyclical set to be correct. So, unlike the Kiesewetter-Lord approach, GI can handle the case of cyclical preferences. Moreover, if my explanation of what makes permissive cases permissive generalizes, then GI should be able to handle *all* permissive cases of the kind Worsnip raises against the Guarantee Hypothesis.

3.3.2. Inter-level Coherence

I'll now turn to inter-level coherence, the other difficult case I want to discuss. As Worsnip argues, it's plausibly a norm of structural rationality that one's first-order beliefs must cohere with one's beliefs about what the evidence supports believing (2021, 81). There are both negative and positive versions of this norm:

> **Negative Inter-Level Coherence**: It is (structurally) rationally prohibited to both believe that your total evidence doesn't sufficiently support believing *p* and believe *p*.
>
> **Positive Inter-Level Coherence:** It is (structurally) rationally prohibited to both believe that your total evidence decisively supports believing *p* and disbelieve or withhold on *p*.

I've formulated the positive version of the norm slightly differently from Worsnip, only to avoid the issues mentioned earlier regarding absences of attitudes, but I don't think it makes an important difference here. Anyway, the above are very plausibly norms of structural rationality. But Worsnip argues that they are counterexamples to the Guarantee Hypothesis on account of the existence of misleading higher-order evidence.

The argument is that misleading higher-order evidence cases involve a conflict between what your first-order evidence supports believing and what your higher-order evidence supports believing about what your first-order evidence supports believing. Consider the following case, originally from Worsnip (2018).[19] Miss Marple is a famous detective who is investigating a murder with her great niece Mable. The two conduct their investigation, but there isn't sufficient evidence regarding who committed the murder. However, in an uncharacteristic error, the normally ultra-reliable Miss Marple declares to Mabel that the evidence supports believing the vicar committed the murder. Worsnip argues that, on account of the misleading higher-order evidence provided by Miss Marple, it would be substantively rational for Mabel to believe the evidence supports believing the vicar committed the murder, but also substantively rational for her withhold on that very proposition.

The problem such cases create for the Guarantee Hypothesis is that they involve two beliefs, each of which is substantively rational (in virtue of being supported by your total evidence), that together form a structurally irrational set. Thus, the Guarantee Hypothesis can't explain the structural irrationality present in misleading higher-order

[19] Cases involving misleading higher-order evidence have also been developed and discussed by Coates (2012), Horowitz (2014), and Lasonen-Aarnio (2020).

evidence cases in terms of guaranteed substantive irrationality. Like permissive cases, such cases are part of a more general class of counterexamples for Worsnip, which he calls *conflict cases*. Because such cases generate a conflict between substantive and structural rationality, the Guarantee Hypothesis can't make sense of them.

Though it's controversial whether such conflicts can arise, I'll accept that they can for the sake of argument.[20] Can GI explain conflict cases? For GI to explain inter-level coherence, it seems it would have to be the case that it isn't possible for both beliefs in the set to be true. But we run into immediate trouble here, because this is obviously not the case. It can both be true that *p* and that your evidence doesn't support believing *p*, and it can both be true that *~p* and that your evidence *does* support believing *p*. Indeed, this is clear independently of whether cases of inter-level incoherence generate conflicts between substantive and structural rationality. So, it seems that GI, far from doing better than the Guarantee Hypothesis, is even less promising for explaining inter-level coherence norms of structural rationality.

Unfortunately, this means GI can't unify all the norms of structural rationality, assuming norms of inter-level coherence are genuine. Fortunately, this doesn't spell disaster for my account. Instead, it illuminates the way forward to find the true account of structural rationality. Here is how. First, consider GI again in full:

> **Guaranteed Incorrectness (GI):** If some set of attitudes S is such that, in virtue of the relations between the constitutive correctness conditions of (at least some of) the attitudes in S, S is guaranteed to contain at least one incorrect attitude, then it is (structurally) rationally prohibited for you to hold S.

[20] For an argument against the possibility of conflict cases, see Neta (2018).

It's plausible that even if it can't explain every norm of structural rationality, GI at least expresses a true conditional. It does seem to be structurally irrational to hold a set of attitudes when at least one of the attitudes is guaranteed to be incorrect.[21]

Next, consider the Guarantee Hypothesis, formulated analogously to GI:

> **Guaranteed Substantive Irrationality (GSI):** If some set of attitudes S is such that, in virtue of the relations between the substantive rationality conditions of (at least some of) the attitudes in S, S is guaranteed to contain at least one substantively irrational attitude, then it is (structurally) rationally prohibited for you to hold S.

Similarly, I think it's plausible that GSI at least expresses a true conditional, even if it can't capture every norm of structural rationality. It does seem to be structurally irrational to hold a set of attitudes when at least one of the attitudes is guaranteed to be substantively irrational. And this fits with what I argued in §2 about substantive rationality being the evidence-relative shadow of correctness. If we should avoid sets of attitudes that guarantee we'll get something incorrect, then we should also avoid sets of attitudes that guarantee we'll be substantively irrational.

What this illuminates is that, insofar as we're trying to get things correct, incorrect attitudes and substantively irrational attitudes each fail *qua* attitude. It's obvious how incorrectness is an attitudinal failure. And it becomes clear how substantive irrationality is an attitudinal failure once we recognize its relationship to correctness, as I argued in §2. Substantive irrationality is an evidence-relative attitudinal failure that's derivative from the fundamental, fact-relative failure of incorrectness. So, if we're trying to get

[21] For considerations of space, I am setting aside worries about cases where a singular belief is guaranteed to be substantively irrational but doesn't seem to be structurally irrational (see Worsnip 2021 for discussion).

things correct, we should avoid sets of attitudes that guarantee attitudinal failure of either kind. This insight points us to the true fundamental norm of structural rationality:

> **Guaranteed Attitudinal Failure (GAF):** If some set of attitudes S is such that, in virtue of the relations between the constitutive correctness conditions of (at least some of) the attitudes in S, S is guaranteed to contain at least one instance of attitudinal failure, then it is (structurally) rationally prohibited for you to hold S.

Unlike previous candidates, GAF clearly can explain the norms of inter-level coherence. Say you both believe that your total evidence doesn't sufficiently support believing *p* and believe *p*. If the former belief is correct, then the latter belief is substantively irrational, and if the latter belief is substantively rational, then the former belief is incorrect. Thus, negative inter-level incoherence involves guaranteed attitudinal failure, so it's structurally irrational. The structural irrationality of positive inter-level incoherence can be explained in much the same way. Moreover, GAF is perfectly compatible with the possibility of conflicts between substantive and structural rationality.

GAF can explain everything that each of GI and GSI can explain, and more. Though I haven't gone through every putative norm of structural rationality, I am optimistic that GAF can provide a unified explanation of all of the norms of structural rationality. Therefore, I propose that GAF is the fundamental norm of structural rationality.

4. Rationality Reunified

With GAF on the table as the fundamental norm of structural rationality, I'm finally in a position to explain how Substantive and Structural Rationality are ultimately unified. Recall that on the view presented in §1.2, the fundamental normative standard for an attitude is its standard of correctness. Fundamentally, our attitudes succeed when they are correct, and fail when they are incorrect. The problem, however, is that we can't

directly ensure that our attitudes are correct, given that the correctness conditions for attitudes implicate the world, and our epistemic grasp of the world is limited. You can't directly ensure that you only believe truths, that you only fear dangerous things, and so on.

This is where rationality comes in. Unlike correctness, rationality is epistemically constrained. The norms of rationality are ones with which we are in a position to comply, even given our epistemic limitations, and which enable us to manifest the commitment to getting things correct. In §2, I explained how norms of substantive rational permission and requirement can be derived from standards of correctness. These norms enjoin us to hold attitudes that are supported by our evidence-relative reasons. And it's easy to see how following these norms manifests the commitment to getting things correct, given our epistemic limitations. For example, we ought to believe in accordance with the evidence, it seems, because evidence is what points us toward the truth. This is the story of how the norms of substantive rationality arise from the fundamental norm of correctness.

Substantive rationality has the label it does because its norms involve assessments of the substantive merits of particular attitudes. But sometimes part of manifesting the commitment to getting things correct (and not getting them incorrect) doesn't involve assessing the substantive merits of particular attitudes. Sometimes we can tell just from the structural relations between attitudes that some set of attitudes is incompatible with manifesting the commitment to getting things correct.

This is exactly how I think GAF, the fundamental norm of structural rationality, arises. GAF prohibits sets of attitudes that involve guaranteed attitudinal failure. The fundamental form of attitudinal failure is incorrectness. That's why I first considered GI, even though it ended up being inadequate. The derivative form of attitudinal failure is

substantive irrationality, and it's a form of attitudinal failure because it is a failure to live up to the attitude's constitutive commitment to getting things correct. Thus, to hold a set of attitudes that guarantees attitudinal failure of either form is incompatible with manifesting the commitment to getting things correct.

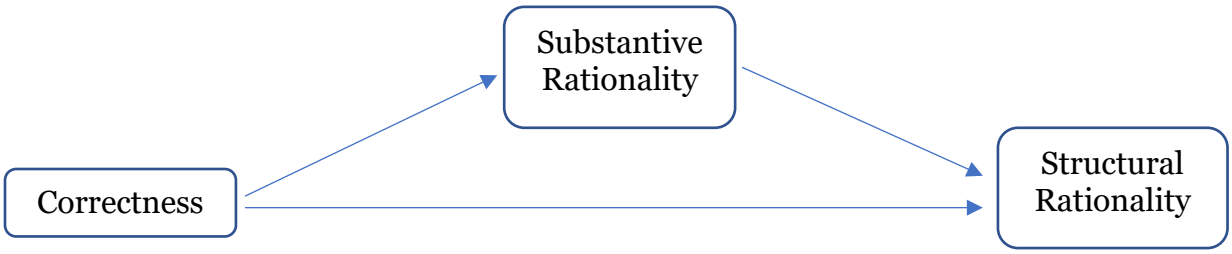
GAF is an indispensable part of the commitment to getting things correct because it allows us to avoid attitudinal failure that's guaranteed in virtue of the constitutive correctness conditions of attitudes. It thereby helps us ensure we get things correct (or at least, don't get things incorrect) without already having substantively assessed each individual attitude for attitudinal failure. Moreover, when we find a set of attitudes to be structurally irrational, this points us to where we *do* need to (re)evaluate individual attitudes for attitudinal failure. Assuming the norms of structural rationality are wide scope, they tell us that we need to reject some attitude in a structurally irrational set, but not which one. They tell us there is attitudinal failure *somewhere*, but they don't tell us where: that's a job for the norms of substantive rationality. Thus, the norms of substantive and structural rationality play complimentary but distinct roles in manifesting the commitment to getting things correct, given our epistemic limitations.

This is how rationality is ultimately unified. While the norms of substantive and structural rationality are distinct, they are both derivative of, and get their normative force from, the fundamental norm of correctness. This account is a form of weak dualism, because it holds there are two kinds of rationality, neither of which can be reduced to or eliminated in favor of the others, but which are ultimately unified. This contrasts with strong dualism, according to which there is no fundamental unity between substantive and structural rationality.

I take it that the kind of disunity that strong dualism posits is, all else equal, theoretically undesirable. More specifically, it's just counterintuitive to hold that there's no single answer to the question "what is rationality?" in virtue of which substantive and structural rationality can be shown to be members of the same kind. This is part of what makes the case for monism appear so strong: it aims to tell us what rationality *is*, without qualification. But given the problems raised for monism, and the motivation this provides for dualism, it can appear that we're stuck between a rock and a hard place. I submit that weak dualism, of the kind I've sketched in this paper, is the way forward. It combines the advantages of monism and strong dualism, while avoiding their shortcomings.

5. Concluding Remarks: Commitment and the Normativity of Rationality

So far, I've argued for a view on which substantive and structural rationality are ultimately unified in terms of correctness. On this view, the fundamental norm on any given rationally evaluable attitude is its standard of correctness. Moreover, each rationally evaluable attitude is rationally evaluable in virtue of having a standard of correctness. The norms of rationality exist because, due to our epistemic limitations, we are unable to directly ensure satisfaction of our attitudes' standards of correctness. In other words, rationality is the epistemically constrained shadow of correctness. Norms of substantive rationality arise solely out of standards of correctness, whereas norms of structural rationality are derivable from a combination of standards of correctness and norms of structural rationality (see figure below).



This is how rationality is reunified. But questions may linger. For one thing, much of what theorists of rationality are concerned with is whether rationality is genuinely normative. Though I've made clear at various points in this paper that I think of rationality as genuinely normative, I haven't said much to argue for this conclusion. I largely focused on establishing that rationality is genuinely normative *if* correctness is genuinely normative, because of how the norms of rationality fall out of standards of correctness. And though I briefly discussed my preferred account of the normativity of correctness in terms of commitment, I endeavored not to rely on it in developing my account of rationality, so as to remain somewhat ecumenical. But there is much more I can say by being less ecumenical, and more committal. That's what I'll briefly gesture at in these last few paragraphs.

Recall how GAF, which I've proposed as the fundamental principle of structural rationality, deals with Worsnip's conflict cases. In such cases, it might be substantively rational to believe p but also substantively rational to believe one's evidence doesn't sufficiently support believing p. Moreover, it can easily turn out *correct* to believe p but also correct to believe one's evidence doesn't sufficiently support believing p. However, there is still guaranteed attitudinal failure in conflict cases: if the latter belief is correct, then the former is substantively irrational, and if the former belief is substantively

rational, then the latter belief is incorrect. Thus, inter-level incoherence violates GAF. And GAF is normative because it is derivable from the fundamental norms of correctness.

However, one might pose a challenge to whether GAF is really normative in conflict cases. If the normativity of rationality really comes down to the normativity of correctness, then why is substantive irrationality genuinely normative in conflict cases where it is possible for all of the attitudes to be correct? In other words, in conflict cases, why not just say that what is rational is to hold all of the correct attitudes, and it doesn't matter that that involves having a substantively irrational attitude? The worry is that, in such cases, if having the substantively rational attitude would lead us away from correctness, substantive rationality is normatively insignificant.[22]

This is a fruitful challenge, because I think its answer sheds further light on the normativity of rationality. However, the answer depends not just on the assumption that correctness is genuinely normative, but on my preferred theory of how it gets to be normative. As I briefly mentioned earlier, I conceive of an attitude as a special kind of intentional mental state that is constituted by a commitment to the attitude's object having a certain property. In committing to the attitude's object having a certain property, the agent subjects that very attitude to the normative standard that the attitude is correct if and only if it has that property. Thus, forming an attitude involves a kind of self-binding, whereby one makes it the case that the standard of correctness for that attitude is genuinely normative. Though I don't have space to defend this account of the normativity of correctness here, I argue for it in other work (Singh 2022, forthcoming).

---

[22] Thanks to David Faraci for raising this challenge and suggesting that I address it.

Even if the commitment-based account explains how correctness is genuinely normative, the challenge is about whether GAF, as a derivative norm, really inherits the normativity of correctness. The worry is essentially that, if the normativity of rationality has to do with the commitment to getting things correct, then in conflict cases (or any other cases where substantive irrationality leads away from correctness), it's unclear whether substantive rationality is really normative. And because the normativity of the incoherence in such cases depends on correctness *and* substantive rationality, it becomes similarly unclear whether *structural* rationality is really normative.

While this would be a serious worry given an instrumentalist understanding of the derivation of norms of rationality, the commitment-based account provides an understanding of the derivation that blunts the force of this worry. If each attitude is constituted by a commitment on the part of its holder to things being a certain way, but we can't directly ensure that we are correct about this, then the only way to manifest this commitment is by following epistemically constrained norms with which we are capable of directly complying. And of course, to have a commitment in the first place, one must be committed to doing what it takes to manifest that commitment. This means that, in holding attitudes, we're committed not just to holding correct attitudes, but also to holding *rational* attitudes when we can't directly ensure our attitudes are correct. The normativity of rationality arises from the normativity of correctness through the commitment-constituted nature of our attitudes.

This account provides the basis for a powerful response to the challenge to GAF's normativity. The deeper explanation of why we should care about substantive irrationality in conflict cases is that, in believing, we commit not just to believing the truth, but also to believing on sufficient evidence. This is because manifesting one's commitment to believe

the truth requires being committed to believing on sufficient evidence. Thus, what appears to be inter-level incoherence in conflict cases is actually incoherence on the same level. If I believe *p* but also believe that my evidence doesn't support believing *p*, then I have at once committed to my evidence supporting believing *p* (in virtue of the former attitude) and committed to my evidence not supporting believing *p* (in virtue of the latter attitude). The reason this is a deeper explanation is that, if we accept my commitment-based account of the normativity of correctness, then it turns out that structural irrationality is simply the holding of inconsistent commitments, and as such, clearly a genuinely normative failure.

This response points the way toward a more ambitious but less ecumenical theory of the normativity of rationality. On such a theory, all genuine normativity arises through the self-binding of commitment. To hold substantively irrational attitudes is a genuine normative failure because it is a failure to live up to the commitments we have made in holding those attitudes. To hold structurally irrational combinations of attitudes is a genuine normative failure because it involves making inconsistent commitments, which guarantees that we will fail to live up to at least some of these commitments. If we opt for this more ambitious theory, the unified account of rationality I've developed in this paper still holds, but it is supplemented by a deeper explanation.

References

Arpaly, N. (2020). Four Notes on John Broome's 'Rationality versus Normativity.' *Australasian Philosophical Review*, *4*(4), 312–320.

Bratman, Michael (2018). *Planning, Time, and Self-Governance: Essays in Practical Rationality*. Oxford University Press.

Broome, John (2007). Does Rationality Consist in Responding Correctly to Reasons? *Journal of Moral Philosophy* 4 (3):349-374.

Broome, John (2020). Rationality versus Normativity. *Australasian Philosophical Review* 4 (4):293-311.

Brunero, John (2020). *Instrumental Rationality: The Normativity of Means-Ends Coherence*. Oxford University Press.

Coates, Allen (2012). Rational Epistemic Akrasia. *American Philosophical Quarterly* 49 (2):113-24.

Danielsson, Sven & Olson, Jonas (2007). Brentano and the Buck-Passers. *Mind* 116 (463):511 - 522.

Fink, Julian (forthcoming). The Essence of Structural Irrationality: The Impossibility of Attitudinal Success. *Journal of Ethics and Social Philosophy*.

Fogal, Daniel & Worsnip, Alex (2021). Which Reasons? Which Rationality? *Ergo: An Open Access Journal of Philosophy* 8.

Fullhart, Samuel & Martinez, Camilo (forthcoming). Coherence as Joint Satisfiability. *Australasian Journal of Philosophy*.

Horowitz, Sophie (2014). Epistemic Akrasia. *Noûs* 48 (4):718-744.

Kiesewetter, Benjamin (2017). *The Normativity of Rationality*. Oxford University Press.

Kiesewetter, B. (2020). Rationality as Reasons-Responsiveness. *Australasian Philosophical Review*, *4*(4), 332–342.

Kolodny, Niko (2005). Why be rational. *Mind* 114 (455):509-563.

Kolodny, Niko (2007). How Does Coherence Matter? *Proceedings of the Aristotelian Society* 107 (1pt3):229 - 263.

Kolodny, Niko (2008). Why Be Disposed to Be Coherent? *Ethics* 118 (3):437-463.

Lasonen-Aarnio, Maria (2020). Enkrasia or evidentialism? Learning to love mismatch. *Philosophical Studies* 177 (3):597-632.

Lord, Errol (2018). *The Importance of Being Rational*. Oxford University Press.

Neta, Ram (2018). Evidence, coherence and epistemic akrasia. *Episteme* 15 (3):313-328.

Scanlon, T.M. (2003). Metaphysics and Morals. Proceedings and Addresses of the American Philosophical Association 77 (2):7-22.

Schoenfield, Miriam (2015). Bridging Rationality and Accuracy. *Journal of Philosophy* 112 (12):633-657.

Schroeder, Mark (2010). Value and the right kind of reason. *Oxford Studies in Metaethics* 5:25-55.

Sharadin, Nathaniel (2016). Reasons Wrong and Right. *Pacific Philosophical Quarterly* 97 (3):371-399.

Singh, K. (2020). Rationality and Kinds of Reasons. *Australasian Philosophical Review*, *4*(4), 386–392

Singh, Keshav (2022). What's in an Aim? *Oxford Studies in Metaethics* 17:138-165.

Singh, Keshav (forthcoming). Belief as Commitment to the Truth. In *What is Belief?* eds. Eric Schwitzgebel and Jonathan Jong. Oxford University Press.

Sylvan, Kurt (2018). Veritism Unswamped. *Mind* 127 (506):381-435.

Sylvan, Kurt L. (2020). An Epistemic Non-Consequentialism. *The Philosophical Review* 129 (1):1-51.

Wedgwood, Ralph (2007). *The Nature of Normativity*. Oxford University Press.

Worsnip, Alex (2018). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research* 96 (1):3-44.

Worsnip, Alex (2021). *Fitting Things Together: Coherence and the Demands of Structural Rationality*. Oxford University Press.