Journal of the American Philosophical Association (2026) 1–18 © The Author(s), 2026. Published by Cambridge University Press on behalf of the American Philosophical Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.
DOI:10.1017/apa.2026.10040

Reasoning, Rationality, and (In)Coherence

ABSTRACT: *Although much ink has been spilled in discussing the relationship between rationality and normative reasons on the one hand and normative reasons and reasoning on the other, surprisingly little has been said about whether we can understand rationality in terms of reasoning. I propose a novel way to understand structural rationality in terms of good reasoning. I argue that structural rationality is best understood in terms of adherence to what I label indefeasible good reasoning. I then offer the beginnings of a theory of substantive rationality in terms of defeasible reasoning. I show how this can provide a unified account of structural and substantive dimensions of rationality, and show that my view also gives the right verdict in some challenging cases for other prominent theories.*

KEYWORDS: Structural Rationality, Good Reasoning, Incoherence, Indefeasible Reasoning, Substantive Rationality

1. Introduction

Despite its significance, there remains considerable debate about the nature of rationality and how to analyze it. One prominent approach to understanding rationality is in terms of reasons-responsiveness (Kiesewetter 2017; Lord 2018). In contrast with this account, other philosophers have argued for an account of normative reasons in terms of good reasoning. (Hieronymi 2005; Hieronymi 2020; Horty 2012; Setiya 2007; Setiya 2014; McHugh and Way 2018; McHugh and Way 2022; Silverstein 2016; Asarnow 2017; Asarnow 2025; Way 2017). This view could perhaps be traced back, rather indirectly, to (Raz 1975; Grice, 2001 [1979]; Williams 1979; Harman 1986).

Quite remarkably and, despite these extensive discussions, the literature has not yet attempted to analyze rationality itself in terms of reasoning. This article aims to fill this gap by offering such an account. In a nutshell, the idea is that *rationality is adhering to good reasoning*. I submit that understanding rationality in terms of reasoning has surprising but welcome consequences. Chief among these is the potential that it offers for a unified account of rationality. Below I focus on giving an account of structural rationality in terms of good reasoning and I will also show how this account can be extended to substantive rationality.

In §2, I introduce what I mean by good reasoning, and make a distinction between two kinds of reasoning. In §3, I propose some desiderata for an account of structural rationality. In §4, I propose and defend my account of structural rationality in terms of reasoning. Here, I show how this account not only gives the right verdict in well-known cases of structural rationality, it also accommodates the proposed desiderata

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

and some crucial intuitions that remain elusive for some other accounts. I respond to objections in §5, followed by brief concluding remarks.

2. Good Reasoning

The usual cases of reasoning are transitions from one set of attitudes to a new attitude. Here are some examples:

Crows:

BELIEF I have seen thousands of crows in my life and they were all black
BELIEF All crows are black

Black Swan:

BELIEF I saw a black swan
BELIEF All swans are black

I will call the attitudes that the agents start with the *premises* and what they end up with the *conclusion*. In these examples, I first indicated the type of each attitude and then its content. A pattern of reasoning can have multiple premises but only one conclusion. What I mean by a 'pattern' of reasoning is more fine-grained than standard categories such as deductive or inductive reasoning. Metaphysically speaking, we can think of a pattern of reasoning as an ordered pair of a set of premises and a conclusion. Each premise or conclusion is an attitude with its respective content. So each of these examples can be seen as a pattern of reasoning (McHugh and Way 2022, Chapter 2). Note that a pattern of reasoning can be comprised of attitudes other than beliefs (Drucker 2021):

Driver:

BELIEF I am driving on a highway with 60 mph speed limit
INTENTION To not go above 60 mph

Magnetic Monopoles:

BELIEF The first expert is saying that magnetic monopoles exist
BELIEF The second expert is saying that magnetic monopoles do not exist
BELIEF Physics is not my forte
SUSPENDING BELIEF Magnetic monopoles exist

In the former, an intention is formed based on a belief. In the latter, although a certain content is *considered*, forming belief with that content (or its negation) is withheld.

Some patterns of reasoning are good, while others are not. The Black Swan example manifests a bad pattern of reasoning. What constitutes good reasoning has been the subject of many debates. But I want to emphasize that to say that a pattern is good, is to say something about the transition between attitudes, rather than something about the *correctness* or *fittingness* of the attitudes you begin with. (This enables us to accommodate the *subjective* dimension of rationality more

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

easily.) For the purposes of this article, our intuitive grip on which patterns of reasoning are good and which ones are not is sufficient. (See McHugh and Way 2018 for an account of good reasoning.)

A key feature of reasoning is that it can get defeated. This means that a good pattern of reasoning might turn into a bad one with further added premises. Take the following famous example from (Williams 1979, 102):

Gin and Tonic:

BELIEF The stuff in the glass looks like gin and tonic and smells like it too
BELIEF The stuff in the glass is gin and tonic

This is a good pattern of reasoning. Now, suppose that the bartender says that she has poured petrol in the glass:

Petrol:

BELIEF The bartender says that she has poured petrol in the glass
BELIEF The stuff in the glass looks like gin and tonic and smells like it too
BELIEF The stuff in the glass is gin and tonic

This pattern, however, is not good. So we can say that the original pattern was defeated with this new premise.

A distinction that becomes crucial in the following is between *indefeasible* and *defeasible* patterns of reasoning. The epitome of indefeasible patterns of reasoning are those that use logical laws like *modus ponens*:

Global Warming:

BELIEF Politicians don't care about global warming
BELIEF If politicians don't care about global warming then humanity will face grave consequences
BELIEF Humanity will face grave consequences

In order to understand which patterns of reasoning are indefeasible, we need to take a look at the ways that they can be defeated. The three widely recognized ways that a pattern of reasoning could be defeated are via rebutting defeaters, undercutting defeaters, and rebutting reason-defeating defeaters (Pollock 1986; Bergmann 1997). In what follows, I will not discuss higher-order defeaters and focus solely on these three types. In both rebutters and undercutters, the premises are assumed to be correct. In the case of rebutters, though the justificatory power of the premises is presupposed, the conclusion is claimed to be incorrect. In the case of undercutters, the very justificatory power of the premise is challenged. However, in reason-defeating defeaters, one of the premises of the pattern of reasoning is shown to be incorrect.

The difference between these three types of defeat could be illustrated with the help of an example: Suppose Henry sees several red barns in a particular area. On the basis of this, he concludes that all the barns in that area are red. His reasoning could be defeated in several different ways: First, it could be that nearly all of the barns in that area are actually red, except for one that is not in Henry's visual field (rebutting defeater). Second, Henry's observation might lack sufficient justificatory power,

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

meaning that, for instance, he has seen only a few of the thousands of barns in the area so he should not have concluded that all the barns are red (undercutting defeater). Third, he might be mistaken in the premise of his reasoning. It could be that the barns that he has seen are not in fact red, but, due to some special optical illusion, they merely seemed red to him (rebutting reason-defeating).

Indefeasible patterns of reasoning can be defeated exclusively by rebutting reason-defeating defeaters. For example, in Global Warming, the pattern of reasoning is defeated only if either of the premises is incorrect. If all of its premises are correct, the conclusion would *inevitably follow*. So a pattern of reasoning is indefeasible if and only if it can be defeated only via rebutting reason-defeating defeaters. I'll call a pattern of reasoning defeasible if it is not indefeasible.[1]

As mentioned in the previous section, many philosophers have proposed accounts of normative reasons in terms of good reasoning. As explored in the upcoming section, many have also thought that there must be a strong relationship between rationality and normative reasons. So, it's fitting to seek a story of rationality told in terms of good reasoning. To motivate the need for this view, I now turn to some desiderata for a theory of structural rationality.

3. Two and a Half Desiderata

In this section, I will set forth two and a half desiderata for a theory of structural rationality. Although I'm skeptical that current accounts of structural rationality are capable of fulfilling these desiderata, I will leave this negative task for another time. Instead, more positively, I will highlight that satisfying these desiderata would be a clear virtue for an account of structural rationality that is told in terms of reasoning.

3.1. Unification

To lay out the first desideratum, we should discuss a key distinction in the rationality literature. *Substantive* rationality, in a slogan, is responding rightly to one's reasons. On the other hand, an agent is *structurally* rational so long as her attitudes fit together well and are without incoherency. Once it was thought that rationality was exhausted by substantive rationality. (Williams 1979) However, there are currently various ways of understanding the relationship between substantive and structural rationality. Monists still insist that only one of these notions is enough to explain rationality. (Kiesewetter 2017; Lord 2018) Dualists, conversely, maintain that neither of these two notions of rationality could be reduced to or eliminated by the other. (Worsnip 2021)

[1] Accordingly, it is crucial to distinguish between an agent's awareness of a defeater and the defeat of a pattern of reasoning itself. What matters regarding the latter is whether the transition from a set of premises to a conclusion remains a good pattern of reasoning once new premises are introduced. As (McHugh 2024) puts it, the question concerns the goodness of the reasoning itself, not the agent's epistemic access to the new premise. Thus, defeat pertains to the objective status of the reasoning pattern, not the agent's subjective access to potential defeaters. For further discussion, see (McHugh 2024). I'm grateful to an anonymous reviewer for prompting this clarification.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

Recently, there have been several attempts to provide accounts of structural rationality in particular. (Worsnip 2021; Lasonen-Aarnio 2021; Fink 2023; Lee 2023; Fullhart and Martinez 2024; Fogal and Risberg 2025) Notwithstanding all the distinct difficulties for each, they fail to provide us with a plausible explanation of why rationality is comprised of two radically diverse phenomena. On these dualist views, it is a mystery as to why we consider both of them types of the same phenomenon and whether there is anything genuinely interesting shared among structural and substantive rationality. Way and Kiesewetter point to a similar problem while discussing Worsnip (2021)'s dualist account:

> Worsnip insists that on his view, structural rationality and substantive rationality are 'equally genuine kinds of rationality'. I wondered what made this so. How is rationality as such to be characterized on Worsnip's view? Without an answer to this question, it's unclear why both structural and substantive rationality count as forms of it. (Way 2022)

> [O]ne would expect … that there is something that structural and substantive rationality have in common, something in virtue of which they are, despite their differences, kinds of one and the same thing. (Kiesewetter 2024, 1232) Cf. (Singh 2025)

On dualist accounts, it's unclear whether there exists anything that unifies rationality at all. Even if there is, the dualist has yet to elucidate it. So the first desideratum is the ability to explain what is shared among structural and substantive rationality.[2]

3.2. Coherence

Consider these pairs:

- Belief
  - {Belief It rains, Belief If it rains, then the streets are wet, Belief The streets are wet}
  - {Belief It rains, Belief 37 is an odd number, Belief There are penguins on all 7 continents}
- Preference
  - {Preference Football to basketball, Preference Basketball to soccer, Preference Football to soccer}
  - {Preference Autumn to spring, Preference Tea to coffee, Preference Pomegranates to cherries}

[2] I thank an anonymous reviewer for informing me of an interesting way of achieving a form of unified view of rationality based on the frameworks developed in (Restall 2005) and (Ripley 2013) and recently put to work in (Brandom and Hlobil 2024). The central idea is to reconceive 'logical consequence' in terms of coherence and rationality. (The same thought could be expanded to also include defeasible reasoning and substantive rationality.)

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

- Intention
  - {Intention to get a cake for your spouse's birthday, Belief It is necessary to go to the bakery in order to get a cake for them, Intention to go to the bakery}
  - {Intention to get a cake for your spouse's birthday, Belief Being human is necessary for Socrates, Intention to cancel your gym membership}

We can easily generate further combinations of attitudes but the point is that none of these sets of attitudes are structurally irrational. However, as Fogal and Risberg (2025) convincingly argue, there is a crucial difference between the former set and the latter set of each of these pairs. While the former *cohere* or *fit* together well, the latter lack coherence; they are in a sense *neutral*.

The literature on structural rationality has mainly focused on cases where a set of attitudes are outright incoherent and has been oblivious to the difference between neutral sets of attitudes and sets of attitudes that cohere together well. What's more, current accounts of structural rationality are unable to explain this difference. And, as Fogal and Risberg (2025) contend, it's reasonable to think that the explanans of incoherence should also be the same as the explanans of neutralness and (strong) coherence.[3] So, following Fogal and Risberg, the second desideratum is that an account that explains what makes a set of attitudes incoherent should also be able to explain neturalness or coherence of a set of attitudes.

3.3. Normativity

If we hear that a law has just been passed requiring that we do not drive our cars faster than 99 percent of the speed of light, we would find it bizarre. We already follow this law, and it would be nearly impossible to break it. In this case, the law would be *anemic* in the sense that it doesn't really ask anything from anyone.

Johnson King (2024) argues that the most developed account of the requirements of structural rationality render it anemic. In essence, these accounts make it nearly impossible to find instances of people violating the requirements of structural rationality. The problem is that in order to explain why a set is incoherent many of the extant accounts appeal to some *constitutive* features or the natures of the attitudes in the set that, in turn, either preclude or make it nearly impossible not to follow the requirements of structural rationality.

To illustrate, consider a hyperbolic theory that claims that a set of attitudes is incoherent if and only if no agent could in principle have all the attitudes in the set. According to such an account, there are no instances of structural irrationality, and, as a result, the requirements of structural rationality are fully idle, as they are always satisfied. This is obviously too extreme. But Johnson King argues that even though

[3] It is evident that there is a sense of 'coherent' that merely means lack of incoherence. Here, I will take no stance on whether our ordinary usage of the term refers to this notion or the stronger one. Cf. (Berker 2024) Rather, regardless of what they should be called, I think these examples show that there is a crucial difference among the non-incoherent sets of attitudes. For further discussion, see (Fogal and Risberg 2025, §§1-2).

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

our current accounts of incoherence do not altogether rule out the possibility of structural irrationality, they make transgressing the requirements of structural rationality nearly impossible.

She argues that according to these accounts, one can only 'appear' to violate the requirements of structural rationality, though one never really breaks them. Johnson King delineates this point while discussing Worsnip's proposal:

> So, the very thing in virtue of which [the agent] appears to transgress her structural-rationality-based reasons turns out to be an instance of exactly the sort of thing that is supposed to be impossible according to the constitutivism underpinning [Worsnip's account of structural rationality] … This is weird. There is a sequence of mental states and processes … that was supposed to be that in virtue of which [she] transgresses the reasons of structural rationality. But the very same sequence of mental states and processes also ensures, given Worsnip's constitutivism, that [her] set of mental states is actually not the incoherent set that it appears to be but merely a similar-looking coherent set. (Johnson King 2024, 10)

Hence, if Worsnip's account (or similar ones) are correct, the requirements of structural rationality become practically inert. Or, as Johnson King puts it, it would be "at best difficult, and at worst impossible, to find instances of people transgressing their structural-rationality-based reasons." (Johnson King 2024, 11) This leaves the supposed norms of structural rationality too weak or anemic.

While Johnson King primarily focuses on Worsnip's account, she extends her critique to several other theories of structural rationality (Johnson King 2024, §4.2). She contends that her argument shows that reformulating the requirements of rationality is also of no help so long as our account of structural rationality appeals to constitutive features of the attitudes. Hence, to preserve the normative dimension of structural rationality, we must seek an account that doesn't rely on constitutive features or the nature of attitudes to explain structural irrationality.

Since the normative dimension of structural rationality is contested (Kiesewetter and Worsnip 2023, §3), I will consider the ability to save the normativity of structural rationality to be half a desideratum. Nonetheless, I will demonstrate how my account remains fully compatible with the normative dimension of structural rationality.

To sum up, our account of structural rationality should be able to explain substantive rationality as well and provide us with a unified explanation of the difference between (strongly) coherent, neutral, and incoherent sets of attitudes. Further, it should not make the (alleged) requirements of structural rationality anemic. In the following section, I will argue for an account that satisfies these desiderata.

4. Rationality in Terms of Reasoning

4.1. Rationality and Good Reasoning

In a slogan, my view is that *rationality is adhering to good reasoning*. To see this proposal in action, we can discuss some of the well-known examples of irrationality:

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

> **Wanda the Wonder Woman:** Wanda's mother usually calls Wanda a 'wonder woman.' Solely based on this, Wanda concludes that she is Wonder Woman. Also, she believes that whoever is Wonder Woman can fly. However, she believes that she cannot fly. Cf. (Fogal 2018)

It seems plausible that there are two things wrong with Wanda's beliefs, not one. First, she is wrong about believing that she is Wonder Woman. The way her mother calls her shouldn't make her believe that she is Wonder Woman. Second, she has a set of incoherent beliefs, whether or not she is Wonder Woman. This fact, that she is irrational in these two ways, should be explained by all theories of rationality.

The explanation in terms of good reasoning seems pretty straightforward: her set of beliefs ({BELIEF she is Wonder Woman, BELIEF Whoever is Wonder Woman can fly, BELIEF she cannot fly}) is inconsistent since each pair of her beliefs can be the premises of a pattern of reasoning that *goes against* the other remaining belief and she is capable of *following* these patterns (I'll come back to these notions shortly). Further, she is irrational in believing that she is Wonder Woman as many good patterns of reasoning go against this belief. In these ways, her irrationality results from her failure to adhere to good reasoning.

Another type of example is having cyclical preferences. Let's imagine that Mike has this set of attitudes: {PREFERRING Kareem to Larry, PREFERRING Larry to Michael, PREFERRING Michael to Kareem}. We can here observe that Mike is structurally irrational in having these attitudes and that there is a good pattern of reasoning from each pair of the attitudes in this set that goes against the other remaining attitude.

Next, suppose that even though Raven has seen thousands of crows in her life which were all black, she believes that there are non-black crows too. This belief of hers is not supported by any piece of evidence. She has formed it out of nowhere, even though she knows of a way to disprove it. Again, her irrationality could be seen as a result of there being a good pattern of reasoning that goes against her belief.

Before proceeding any further, it is important to discuss two concepts that I used in the foregoing analyses. First, 'being capable of a pattern of reasoning' means that the subject possesses the *intellectual capacity* required for that piece of reasoning. So, if the agent followed that pattern, instead of a mere 'mental jump,' this would be an instance of inferring an attitude in the right way.

Second, I used the phrase of one pattern of reasoning 'going against' another attitude. Whenever the conclusion of a pattern of reasoning goes against one's current attitudes, a competing type of response is advised toward the object of that attitude. For example, in doxastic cases, the pattern of reasoning requires the agent to suspend believing or believe the negation of what they currently believe. Or, if the conclusion of a pattern of reasoning is INTENTION to φ and the agent lacks that intention, this is also a case where a pattern of reasoning goes against the agent's set of attitudes. The same goes for other types of attitudes.

One final point. As astutely noted by Harman (1986, 12), we need not "clutter our minds." We don't need to believe every single consequence of our beliefs to escape irrationality. A mathematician can be rational even if her set of beliefs doesn't include every single mathematical truth. In what follows, I take it that no agent can be

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

charged with irrationality if she lacks a certain attitude unless that attitude is, in a certain sense, *considered* by her or she has *given a verdict* on it. (Of course, if she has been irresponsible in failing to consider that attitude, she is criticizable for that.) This will be relevant especially when I discuss substantive rationality. But let's first turn to structural rationality.

4.2. Structural Rationality

Now, we are in a position to give an account of what it is for a set of attitudes to be incoherent, and to likewise give an account of structural irrationality. Worsnip highlights the point that there are different ways to be incoherent:

> [T]he instances of incoherence are a diverse bunch, including a range of different phenomena. Some of these instances involve combinations of doxastic states, some involve combinations of practical states, and some involve combinations of doxastic and practical states. In light of this, it's reasonable to wonder what, if anything, these instances have in common and whether there is any kind of general method for identifying which combinations of states really are incoherent. (Worsnip 2021, 128)

There is a straightforward explanation of this in terms indefeasible reasoning:

> **Structural (Ir)rationality in Terms of Reasoning:** A set of attitudes is incoherent if and only if (and because) there is an indefeasible pattern of reasoning from a group of them that goes against another one of the attitudes in the set.

This is observable in the established categories of structural irrationality. The most recognized category of incoherency is that of having inconsistent beliefs. As we saw in the Wonder Woman case, here is one of the forms this can take: {BELIEF p, BELIEF if p then q, BELIEF $\neg q$}. Here, there is an indefeasible pattern of reasoning from each pair of the attitudes that goes against the other one in the set. The same goes for all inconsistent beliefs.[4] Cases of having cyclical preferences are also similar to this.

The other widely discussed case of structural irrationality is means-end incoherence. It often takes the following form: {INTENTION φ, BELIEF It is necessary to ψ in order to φ, NOT INTENDING ψ} Once again, there is an indefeasible pattern of reasoning from these attitudes:

INTENTION φ
BELIEF It is necessary to ψ in order to φ
INTENTION ψ

[4] The literature is divided on whether a singleton set with a contradictory belief is structurally irrational or not. For contrasting views, see (Broome 2013, 153) and (Fink 2023, 402) (and references therein). Analogous to how certain tautologies can be derived via deductive inference that begins with no premises, we might also acknowledge patterns of good reasoning that start with no premises which their conclusion goes against contradictory beliefs. This would allow the current framework to incorporate such cases. Thanks to an anonymous reviewer for raising this point.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

And the conclusion of this reasoning goes against the latter state in the set above.[5] However, as pointed out in Kolodny (2005, 518ff), here there aren't indefeasible patterns of reasoning that go against *all* of the attitudes present in the incoherent set, unlike what we saw in the case of inconsistent beliefs. For example, the following pattern of reasoning could be deemed to be a case of bootstrapping:

INTENTION φ
NOT INTENDING ψ

BELIEF It is not necessary to ψ in order to φ

So, in all structurally irrational sets, there is an indefeasible pattern of reasoning going against at least one attitude.

Turning to akrasia, ({BELIEF All things considered, I ought to φ, NOT INTENDING φ) we can have a straightforward explanation of why this is also an instance of incoherency according to our account. The reason is simply that the following is an indefeasible pattern of reasoning:

BELIEF All things considered, I ought to φ

INTENTION φ

Next, let's consider inter-level incoherency:

BELIEF All evidence decisively supports believing p

BELIEF p

I believe that this is an indefeasible pattern of reasoning, just like the other ones above. To see this, we can note that it seems that this pattern cannot be rebutted. For anything that might seem like a rebuttal would make the other premise (BELIEF My total evidence decisively supports believing p) incorrect and the defeater would need to be a rebutting reason-defeating defeater instead. Recall that in the usual cases of rebuttal, the correctness of other premises is not altered with the further premise. The same holds for undercutters. Hence, it is reasonable to assume that the pattern is indefeasible. The same considerations apply to other patterns of reasoning that I claim to be indefeasible and we can see that they can get defeated only via rebutting reason-defeating defeaters.

Above, I have demonstrated how inconsistent beliefs, means-end incoherence, cyclical preferences, akrasia, and inter-level incoherency are categorized as structurally irrational according to my account. Here, I want to stress how my view can also give the right verdict on types of incoherency that hitherto have been

[5] As noted earlier, one doesn't become irrational if she has never even *considered* some of the consequences of her beliefs. What about intentions? Do we become irrational if we do not form intentions for every action that is necessary to realize our intended ends? Similarly, I don't think so. We form numerous intentions in the course of a day, and often have beliefs about what is required to realize them, many of which are implemented automatically. Thus, we can remain rational even if not all such means are explicitly intended. So what can constitute structural irrationality is when we've *considered* an attitude but opted not to form it. Cf. (Millar 2004, 73-77) So, in what follows, when I talk about NOT INTENDING what I mean is more than a mere and utter lack of an attitude; it is that the agent has *considered* but refused to form it. Thanks to an anonymous reviewer for inviting me to make this clearer.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

neglected in the literature. These include (but are not limited to) having inconsistent credences, hopes, desires, admirations, loathes, and fears. Here is a simple example of one such case: {BELIEF All things considered, Mandela is admirable, NOT ADMIRING Mandela} Here, there is an indefeasible pattern of reasoning that goes against the latter attitude in the set:

BELIEF All things considered, Mandela is admirable
ADMIRING Mandela

Notice the striking similarity with cases of akrasia. Some might try to argue that this is not a genuine example of structural (ir)rationality. Considering the scarcity of this type of example in the literature this sentiment isn't far-fetched. I hold that anyone who believes that this example is not a bona fide type of structural irrationality has to also concede that the pattern of reasoning corresponding to it is not indefeasible. All (purported) explanations of why this set of attitudes is not structurally irrational simultaneously provide (purported) rebutting or undercutting defeaters for the aforementioned pattern of reasoning, thereby making the pattern defeasible. In other words, there is (perhaps implicitly) at least one rebutting or undercutting defeater in whatever story told in order to argue that this set of attitudes is not structurally irrational. This, again, shows that structural rationality and indefeasible patterns of reasoning go hand in hand.

Interestingly, the literature on emotions has addressed similar cases and the fact that they can behave like other instances of incoherency (D'Arms and Jacobson 2003; Grzankowski 2020), but such cases have not made their way into debates about structural rationality. To be clear, here I do not want to unequivocally claim that such cases should be categorized as instances of structural irrationality and to tie the fate of my account to this idea. Rather, the claim is that the account is capable of generating new and hopefully fruitful discussions about structural irrationality.

4.3. Substantive Rationality

I will now show that the account of structural rationality in terms of good reasoning can naturally be extended to substantive rationality as well. A full defense would require much more space than what is available here but I will demonstrate why this approach is promising. The idea is that *our best patterns of reasoning are what guarantee the rationality of our attitudes*. By following them, we ensure that nothing has gone awry.[6] And whenever we are being irrational in holding one of our attitudes, there seems to be a good pattern of reasoning that

[6] A quick note about the reverse order of explanation that could be attributed to (Broome 2013). If the positive account of this article is successful, it shows that rationality is only the bare minimum of what good reasoning promises us to achieve if we adhere to its rules. Even a subject with only a handful of coherent and justified beliefs is still 'certified' as rational. For some arguments why we cannot analyze good reasoning in terms of rationality, see (McHugh 2024, §4) (that it fails to capture the defeasible nature of good reasoning) and (Worsnip 2018) and (Worsnip 2019) (that it faces preface paradox-style problems and other issues).

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

shows where things have gone wrong. So, a natural proposal would be something along these lines:

> **Substantive (Ir)rationality in Terms of Reasoning (1):** Having an attitude is irrational (substantively) if and only if (and because) one of the patterns of good defeasible reasoning that the agent is capable of goes against that attitude.

But this won't do. For the pattern of reasoning that is not followed might be one that is already defeated. Recall the Petrol case. Bernie is looking at a glass in front of him that looks like gin and tonic and smells like it too. However, the pattern of reasoning to the conclusion that it is gin and tonic gets defeated by the information he receives. Even though one of Bernie's patterns of reasoning requires him to believe that it is gin and tonic, his *more encompassing* pattern of reasoning defeats it. Bernie would be irrational unless he accepts the conclusion of this latter pattern of reasoning. Therefore, we should modify the account as follows:

> **Substantive (Ir)rationality in Terms of Reasoning (2):** Having an attitude is irrational (substantively) if and only if (and because) among the defeasible good patterns of reasoning that the agent is capable of, there is an encompassing one that goes against that attitude.

Being more encompassing than another pattern of reasoning means that the former pattern of reasoning has all the premises of the latter and also other premises that can affect the conclusion.[7]

One of the divisive issues about substantive rationality is the question of which *type* of reasons is relevant in assessing an agent's rationality. Following Fogal and Worsnip, I think it is necessary to distinguish three types of reasons: *fact-relative*, *belief-relative*, and *evidence-relative*. (Fogal and Worsnip 2021; Worsnip 2021)

Simply put, a fact-relative reason is an objective reason that might be completely out of the agent's epistemic reach. On the other hand, a belief-relative reason is a subjective (or apparent) reason that is part of the agent's set of beliefs. Finally, an evidence-relative reason is, as it were, at the halfway point of these two other notions. Unlike fact-relative reasons, evidence-relative reasons have a certain sort of epistemic relationship with the agent. But, unlike belief-relative reasons, they need not necessarily be one of the agent's beliefs. For example, in some cases one might have evidence for a belief but still refuse to believe it, say, because of wishful thinking.

It is easy to see why fact-relative reasons are not apposite for evaluating the rationality of agents. If an agent lacks any epistemic connection with a reason, without being intentionally negligent, she would not be considered irrational for

[7] I think that this notion can also be used to account for the alleged permissive cases, where conflicting beliefs (say, q and $\neg q$) are rationally compatible with a set of beliefs p. If one only believes p, she is rationality permitted to believe in either q or $\neg q$. But once she believes in one of these two, she will be capable of a more encompassing pattern of reasoning that goes against believing its negation. Thanks to an anonymous reviewer for this point.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

not heeding to that fact. There's something internal and subjective about rationality that fact-relative reasons cannot capture.

That leaves us with the belief-relative view and the evidence-relative view. Here, I will take no stance on which one of these notions is apt for evaluating substantive rationality. But I believe that the reasoning view of rationality is capable of accommodating both these views. Above, I tacitly assumed the belief-relative view. So, here is a more explicit formulation of the belief-relative view:

> **Belief-relative Substantive (Ir)rationality in Terms of Reasoning:** Having an attitude is irrational (substantively) if and only if (and because) among the defeasible good patterns of reasoning that the agent is capable of, there is an encompassing one that goes against that attitude and the agent has all the premise attitudes.

Let's turn to the evidence-relative view. To stop the evidence-relative view from collapsing into the belief-relative view, we should grant the existence of non-doxastic evidence, as in cases like intuition or perceptual experience. An agent could be irrational if she does not heed this type of evidence. For example, If Ellie is seeing a gray elephant, but, for no good reason, she believes that the elephant in front of her is purple, she is being irrational.

To get an evidence-relative view, the belief-relative account can be revised to account for irrationality in such scenarios. Even if Ellie is refraining from believing that the elephant is gray, she is in an *epistemic position to know* that 'it seems to me that the elephant in front of me is gray.'[8] And that could be a premise in a pattern of reasoning that goes against her belief that the elephant is purple. More generally, in cases like this, the agent is in a position to know the fact that 'there is evidence for *p*' which could be a premise for a defeasible pattern of reasoning that goes against her irrational attitude. Therefore, to accommodate the evidence-relative view, we only need to revise the belief-relative formulation to also include the possible premises that the agent is in a position to know:

> **Evidence-relative Substantive (Ir)rationality in Terms of Reasoning:** Having an attitude is irrational (substantively) if and only if (and because) among the good patterns of reasoning that the agent is capable of, there is an encompassing one that goes against that attitude and the agent is in an epistemic position to know all the premise attitudes.

And if all the premises of this pattern are among the agent's set of attitudes, then the pattern is defeasible (since if it had been indefeasible, then she would have been

[8] For a discussion of this notion see (Lord 2018, Chapter 3). Cf. (Kiesewetter 2017, Chapter 8). As I employ this notion in order to capture cases where an agent fails (or refuses) to believe her evidence, this notion is limited, first, to beliefs and, second, to beliefs that are both true and justified. Thus, by the set of attitudes an agent is in an epistemic position to know, I mean the agent's current attitudes together with the beliefs she is in a position to know. (Note that the core idea behind the evidence-relative view is that this set can determine the irrationality of all the attitudes, including the non-belief ones, and this is captured here by appealing to encompassing patterns of reasoning.)

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

structurally irrational and not substantively so) and if some of the premises are not among her attitudes, then the pattern can be either indefeasible or defeasible.

So the Evidence-relative version not only captures cases from the Belief-relative version, but also accounts for cases where evidence is not part of the agent's set of beliefs. Hence, this shows that Ellie is being irrational because she is in an epistemic position to know all the premises of a defeasible good pattern of reasoning that goes against her current belief.[9]

4.4. Desiderata

We can now return to the desiderata. Recall the first desideratum: that an account of rationality should offer a unified explanation of structural and substantive rationality. My account is a monist one that reduces and explains the substantive-structural distinction in rationality in terms of the indefeasible-defeasible distinction in patterns of reasoning.[10] One worry might be that this type of reduction and explanation seems like pushing the bump around the rug; now we are facing distinct types of good reasoning. However, this account reduces and explains the distinction in terms of a fundamental demarcation: an opposition between patterns of reasoning that have the property of being indefeasible and those that lack it. So it seems like a natural place for the explanation to end.

Recall that the second desideratum was that an account of structural rationality should explain the difference between these sets:

- {Belief It rains, Belief If it rains, then the streets are wet, Belief The streets are wet}
- {Belief It rains, Belief 37 is an odd number, Belief There are penguins on all 7 continents}

While the attitudes in the former set cohere together well, neutral sets like the latter lack coherence. According to the account of rationality in terms of reasoning, what makes a set incoherent is the existence of an indefeasible pattern of reasoning that goes against one of the attitudes in the set. So, in the opposite direction, a set of attitudes coheres well when there is an indefeasible pattern of reasoning with its

[9] An anonymous reviewer aptly noted that while we tend to describe a set of attitudes as structurally irrational, in cases of substantive irrationality we typically identify a particular attitude as the source of the irrationality. This, I suggest, may be further reason to favor the evidence-relative view, which offers a principled way to determine which attitude should be revised when multiple candidates are involved. (This also bears on the debate over whether rational requirements are wide-scoped or narrow-scoped.) However, since I aim to remain neutral on this issue in the present article, I do not pursue it further.

[10] An anonymous reviewer rightly pointed out that structural and substantive irrationality still differ in ways that go beyond the distinction between indefeasible and defeasible reasoning, for example one is typically ascribed to a set of beliefs but the other is often about a single attitude. While I acknowledge these differences, I believe that the central thesis of this framework still succeeds in capturing the shared underlying nature of both forms of irrationality—namely, that each arises from a failure to respond appropriately to different kinds of reasoning. Even though they have their own distinctive features, their nature is still *explained* in a unified manner (as stemming from failures of appreciating indefeasible and defeasible reasoning).

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

conclusion being part of that set (and there are no indefeasible patterns of reasoning that go against any of the attitudes in the set to make it incoherent). Further, a set of attitudes is neutral (like the latter one here) whenever, from the attitudes in the set, there aren't any indefeasible patterns of reasoning neither going against nor for the other attitudes in the set. So we can see how neatly a set of attitudes being incoherent, neutral, or coherent is explained by facts about indefeasible patterns of reasoning. Based on this account, we can even start discussing different 'degrees' of coherence and how a set can be 'more' coherent than another one based on the number of indefeasible (or, in a lower degree, defeasible) patterns of reasoning that exist for the attitudes in each set.

The final desideratum was that it should not render the (alleged) norms of structural rationality anemic; it shouldn't be nearly impossible to find people with an incoherent set of attitudes that are transgressing the norms of structural rationality. The account developed here successfully meets this criterion. Recall that Johnson King (2024) argued that as other accounts of structural rationality appeal to constitutive aims or dispositions, they cannot accommodate this normative dimension. Their way of tying what constitutes structural rationality with what constitutes attitudes made it nearly impossible to have such attitudes while genuinely transgressing the normative requirements of structural rationality. Hence, we were left with a nearly idle or anemic normative dimension for structural rationality. But, contrary to other accounts, the theory presented here does not rely on constitutive features of attitudes to explain structural rationality. Instead, it employs indefeasible reasoning to analyze incoherency. Crucially, adhering to good reasoning is not a constituent of having an attitude. So this doesn't threaten the supposed norms of structural rationality. One can have an incoherent set of attitudes and yet simply decide not to follow good patterns of reasoning to revise her attitudes. Therefore, my account ensures the possibility of people being incoherent and at the same time, infringing the requirements of structural irrationality.

As a bonus, I suggest that this account can unlock a new avenue for discussing whether the norms of structural rationality are wide- or narrow-scoped. If we remember the means-end incoherency, we can hold that whether the requirements of structural rationality are wide- or narrow-scoped is connected with whether multiple patterns of good reasoning are violated or only one is violated. If the former, the requirement should be taken to be wide-scoped; else the requirement is narrow-scoped. This is only a first step but it shows that my account fully accommodates the (alleged) normative dimension of structural rationality and can start new conversations about them.

5. Objections and Replies

5.1. All Rational Attitudes Come from Reasoning?

The account above is supposed to be a *metaphysical* analysis of rationality and does not imply that all rational attitudes come from reasoning. It is undoubtedly true that many of our rational attitudes do not come from any sort of deliberation. What the account here is committed to is that whenever an attitude is irrational, there is a good pattern of reasoning that goes against it.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

5.2. Reason without Reasoning?

Let's now consider so-called 'elusive reasons.' For example:

> **Surprise Party:** [C]onsider the case of Nate, who hates all parties except for successful surprise parties thrown in his honor. Given Nate's situation, the fact that [there is a surprise party waiting for him now at home] is a reason for him to go home. But it isn't a reason that Nate could know about or act on. (Schroeder 2007, 33)

In this example, it is argued that there is a normative reason for Nate to go home, but there is no 'reasoning path' for him to be motivated to do so (see also Markovits 2014). As the account of substantive rationality developed here hinges on the role that normative reasons can play in good reasoning, it might be thought that such cases undermine this view.

Crucially, by definition, these elusive normative reasons are always fact-relative and beyond the epistemic reach of subjects. Therefore, they are not relevant in determining whether or not an agent is rational. If Nate doesn't decide to go home, although there is a normative reason that he is not following, that doesn't make him irrational. This shows that the account gives the desired verdict in these cases. See also (McKeever and Ridge 2012; Sinclair 2016; Way and Whiting 2016; Paakkunainen 2017; Kietzmann 2023, §4; Asarnow 2025).

5.3. Means-end Reasoning Is not Indefeasible

I have maintained that means-end reasoning is an example of indefeasible reasoning. For example, let's say that you intend to keep your clothes clean, and you believe that to do so, it's necessary to intend to not jump into the nearby pond. However, you may see a child drowning and thus come to intend to jump into the pond in order to save the kid. Does this way of defeating the means-end pattern of reasoning show that this type of reasoning is not indefeasible?

To show that means-end patterns of reasoning are not indefeasible, the objector has to provide an example where they can be defeated with either a rebutting defeater or an undercutting defeater. However, in the example given here, the new premise does the job of a rebutting reason-defeating defeater. It alters the correctness of the other premise (the intention to keep your clothes clean). This is how a rebutting reason-defeating defeater works. In this regard, there is nothing special about means-end reasoning since, for example, patterns that use modus ponens also have these defeaters. Generally, cases purporting to undermine the indefeasibility of such patterns are cases of attitudes with concealed conflicting fittingness conditions. (McHugh and Way 2022, §3.5.3)

6. Conclusion

I have argued for an account of structural rationality in terms of good reasoning. After defining indefeasible and defeasible patterns of reasoning, I proposed a set of

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

desiderata for an account of structural rationality. I argued how we can analyze structural (ir)rationality in terms of indefeasible patterns of reasoning and substantive (ir)rationality in terms of defeasible patterns of reasoning. I then demonstrated how this account explains the unity of rationality, the distinction between incoherent, neutral, and coherent sets of attitudes, and, finally, preservers the alleged normative dimension of structural rationality. Further, it can also open up a new way of thinking about its requirements. Hence, it uncovers a novel way of thinking about rationality in general.[11]

HOSSEIN AYATI
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES, TEHRAN, IRAN
seyedhosseinayati@gmail.com

References

Asarnow, Samuel. 2017. "The Reasoning View and Defeasible Practical Reasoning." *Philosophy and Phenomenological Research* 95 (3): 614–636.
Asarnow, Samuel. 2025. *Reasons Last: Agency, Morality, and the Reasoning View*. Oxford University Press, February.
Bergmann, Michael. 1997. "Internalism, Externalism and the No-Defeater Condition." *Synthese* 110 (3): 399–417.
Berker, Selim. 2024. "Is There Anti-Fittingness?" *Ergo: An Open Access Journal of Philosophy* 11 (39):1051–1082.
Brandom, Robert, and Ulf Hlobil. 2024. *Reasons for Logic, Logic for Reasons: Pragmatics, Semantics, and Conceptual Roles*. Edited by Robert Brandom. New York.
Broome, John. 2007. "Does Rationality Consist in Responding Correctly to Reasons?" *Journal of Moral Philosophy* 4 (3): 349–374.
Broome, John. 2013. *Rationality Through Reasoning*. Malden, MA: Wiley-Blackwell.
D'Arms, Justin, and Daniel Jacobson. 2003. "Viii. The Significance of Recalcitrant Emotion." *Royal Institute of Philosophy Supplement* 52:127–145.
Drucker, Daniel. 2021. "Reasoning Beyond Belief Acquisition." *Noûs* 56 (2): 416–442.
Fink, Julian. 2023. "The Essence of Structural Irrationality." *Journal of Ethics and Social Philosophy* 26 (2).
Fogal, Daniel. 2018. "On the Scope, Jurisdiction, and Application of Rationality and the Law." *Problema* 12:21–57.
Fogal, Daniel & Risberg, Olle (2025). "Coherence and Incoherence." *Philosophical Review* 134 (4): 405–454.
Fogal, Daniel, and Alex Worsnip. 2021. "Which Reasons? Which Rationality?" *Ergo: An Open Access Journal of Philosophy* 8.
Fullhart, Samuel & Martinez, Camilo. 2024. "Coherence as Joint Satisfiability". *Australasian Journal of Philosophy* 102 (2):312–332.
Grice, Paul. 2001[1979]. *Aspects of Reason*. Oxford, GB: Clarendon Press.
Grzankowski, Alex. 2020. "Navigating Recalcitrant Emotions." *Journal of Philosophy* 117 (9): 501–519.
Harman, Gilbert. 1986. *Change in View: Principles of Reasoning*. Cambridge, MA, USA: MIT Press.
Hieronymi, Pamela. 2005. "The Wrong Kind of Reason." *Journal of Philosophy* 102 (9): 437–457.
Hieronymi, Pamela. 2020. "Reasoning First." In *The Routledge Handbook of Practical Reasoning*, edited by Ruth Chang and Kurt Sylvan.

[11] I am grateful to the editor and two anonymous reviewers for their feedback. I would also like to thank Alireza Kazemi, Conor McHugh, Jonas Olson, Olle Risberg, Jonathan Way, and the participants of the Higher Seminar in Practical Philosophy at Stockholm University for their thoughtful comments and suggestions.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press

Horty, John. 2012. *Reasons as Defaults*. Oxford, England: Oxford University Press USA.
Johnson King, Zoë. 2024. "A Puzzle About the Unity and Normativity of Structural Rationality." *Erkenntnis* 90 (7):3169–3194.
Kiesewetter, Benjamin. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.
Kiesewetter, Benjamin. 2024. "Review of Fitting Things Together: Coherence and the Demands of Structural Rationality." *Mind* 133 (532):1229–1238.
Kiesewetter, Benjamin, and Alex Worsnip. 2023. "*Structural Rationality*." Stanford Encyclopedia of Philosophy.
Kietzmann, Christian. 2023. "Why Reasons and Reasoning Don't Come Apart." *Synthese* 202 (5): 1–15.
Kolodny, Niko. 2005. "Why Be Rational." *Mind* 114 (455): 509–563.
Lasonen-Aarnio, Maria. 2021. "Coherence as Competence." *Episteme* 18 (3): 353–376.
Lee, Wooram. 2023. "What is Structural Rationality?" *The Philosophical Quarterly* (July): pqad072.
Lord, Errol. 2018. *The Importance of Being Rational*. Oxford, UK: Oxford University Press.
Markovits, Julia. 2014. *Moral Reason*. Oxford, GB: Oxford University Press.
McHugh, Conor. 2024. "Norms of Reasoning." *Philosophy Compass* 19 (7): e13008.
McHugh, Conor, and Jonathan Way. 2018. "What is Good Reasoning?" *Philosophy and Phenomenological Research*, 153–174.
McHugh, Conor, and Jonathan Way. 2022. *Getting Things Right: Fittingness, Value, and Reasons*. Oxford: Oxford University Press.
McKeever, Sean, and Michael Ridge. 2012. "Elusive Reasons 1." *Oxford Studies in Metaethics* 7.
Millar, Alan. 2004. *Understanding People: Normativity and Rationalizing Explanation*. Oxford University Press UK.
Paakkunainen, Hille. 2017. "Can There Be Government House Reasons for Action?" *Journal of Ethics and Social Philosophy* 12 (1): 56–93.
Pollock, John L. 1986. *Contemporary Theories of Knowledge*. London: Hutchinson.
Raz, Joseph. 1975. *Practical Reason and Norms*. London: Hutchinson.
Restall, Greg. 2005. "Multiple Conclusions." In *Logic, Methodology, and Philosophy of Science*, edited by Petr Hájek, Luis Valdés-Villanueva, and Dag Westerståhl. College Publications.
Ripley, David. 2013. "Paradoxes and Failures of Cut." *Australasian Journal of Philosophy* 91 (1): 139–164.
Schroeder, Mark. 2007. *Slaves of the Passions*. New York: Oxford University Press.
Setiya, Kieran. 2007. *Reasons Without Rationalism*. Princeton University Press.
Setiya, Kieran. 2014. "What is a Reason to Act?" *Philosophical Studies* 167 (2): 221–235.
Silverstein, Matthew. 2016. "Reducing Reasons." *Journal of Ethics and Social Philosophy* 10 (1): 1–22.
Sinclair, Neil. 2016. "On the Connection Between Normative Reasons and the Possibility of Acting for Those Reasons." *Ethical Theory and Moral Practice* 19 (5): 1211–1223.
Singh, Keshav. 2025. "Rationality Reunified." *Oxford Studies in Metaethics* 20:175–199.
Way, Jonathan. 2017. "Reasons as Premises of Good Reasoning." *Pacific Philosophical Quarterly* 98 (2).
Way, Jonathan. 2022. "Review of Alex Worsnip, *Fitting Things Together: Coherence and the Demands of Structural Rationality*." In *Notre Dame Philosophy Reviews*.
Way, Jonathan, and Daniel Whiting. 2016. "Reasons and Guidance." *Analytic Philosophy* 57 (3): 214–235.
Williams, Bernard. 1979. "Internal and External Reasons." In *Rational Action*, edited by Ross Harrison, 17–28. Cambridge University Press.
Worsnip, Alex. 2018. "Reasons, Rationality, Reasoning: How Much Pulling-Apart?" *Problema* 12: 59–93.
Worsnip, Alex. 2019. "Isolating Correct Reasoning." In *Reasoning: New Essays on Theoretical and Practical Thinking*, edited by Magdalena Balcerak Jackson and Brendan Balcerak Jackson. Oxford University Press.
Worsnip, Alex. 2021. *Fitting Things Together: Coherence and the Demands of Structural Rationality*. New York: Oxford University Press.

https://doi.org/10.1017/apa.2026.10040 Published online by Cambridge University Press