Review of *Fitting Things Together*

Benjamin Kiesewetter

Forthcoming in: *Mind*, https://doi.org/10.1093/mind/fzad016.

This is a pre-print. Please cite the published version.

Fitting Things Together. Coherence and the Demands of Structural Rationality, by Alex Worsnip. New York: Oxford University Press, 2021. Pp. xvii+335

Two views that have dominated the recent literature on rationality are the coherence-based and the reasons-based conception of rationality. According to the first of these views, rationality is a matter of establishing internal coherence between one's mental states (e.g. Scanlon 2007; Broome 2013). On this conception of rationality, it is an open question whether we ought or have any reason to be rational, and some have argued that this question must receive a negative answer (Kolodny 2005). According to the second view, rationality is most fundamentally a matter of responding to reasons (e.g. Kiesewetter 2017; Lord 2018). Insofar as incoherence is irrational, this is to be explained in terms of a failure to respond to available reasons, and consequently the normative significance of rationality cannot sensibly be questioned.

In *Fitting Things Together*, Alex Worsnip seeks to establish an alternative to both of these views, which he dubs 'dualism about rationality'. According to this third view, there are both structural requirements of rationality, which demand coherence among our attitudes, and substantive requirements of rationality, which demand responsiveness to reasons, and these requirements are irreducible to each other. Dualism agrees with the reasons-based view that there is an important dimension of rationality, the normative significance of which cannot sensibly questioned; but it rejects the thesis that the irrationality of incoherence can be explained in terms of this substantive dimension. Dualism thus agrees with the coherence-based view that there is an important dimension of rationality, the normative significance of which *can* sensibly be questioned, but denies that this structural dimension exhausts what rationality is about.

While Worsnip's dualism thus grants a point to both sides of this dispute, the subtitle of his book already reveals that its focus is clearly on structural rather than substantive rationality.

There is a chapter that outlines Worsnip's preferred conception of substantive rationality (Chapter 2), and another one that argues against attempts to explain reasons in terms of coherence constraints (Chapter 4), but the core aim of the book is to provide a theory of *structural* rationality – more specifically, a theory according to which structural rationality is 'genuine, autonomous, unified, and normatively significant' (ix). In what follows, I will first give a brief overview over the contents and achievements of the book before raising some questions and objections.

1. Summary

The first part of the book defends dualism, the view that 'structural and substantive rationality are two distinct but equally genuine kinds of rationality, neither of which is reducible to the other' (4). The first two chapters defend the distinctness thesis, while the subsequent two chapters focus on the genuineness and the irreducibility theses. The key argument against reducing structural to substantive rationality, elaborated in Chapter 3, is targeted at a crucial assumption on which the proposed reduction relies:

> *The Guarantee Hypothesis:* For any set of attitudinal mental states $\{A_1 \ldots A_n\}$ of the kind associated with structural irrationality, it is guaranteed that at least one of $\{A_1 \ldots A_n\}$ is substantively irrational (i.e. that it is insufficiently supported by the agent's evidence-relative reasons). (54f.)

Worsnip argues that this hypothesis is falsified by certain cases of structural irrationality that do not involve any failure to respond correctly to reasons (such as certain cases of cyclical preferences) and others that can only be avoided by failing to respond correctly to reasons (which may include cases with a preface-paradoxical structure or cases of misleading higher-order evidence).

In the second part of the book, Worsnip develops a positive autonomous theory of structural rationality, which is an important desideratum in the current debate. Chapter 5 provides a unified account of the various instances of structural irrationality, which Worsnip takes to include (among other things) means/end-incoherence as well as inconsistency in belief and incoherence between beliefs and higher-order-beliefs about one's evidence. The account

states that structural irrationality is *co-extensional with* and *grounded in* incoherence, a property that Worsnip in turn explicates by appeal to the following biconditional:

> *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. (133)

So, for example, failing to intend what one believes to be a necessary means to an intended end is structurally irrational on this account because it is incoherent, and its incoherence is implied by (perhaps even amounts to) the (presumed) fact that it is constitutive of intentions that they dispose us to adopt further intentions for means that are believed necessary, insofar as our means/end-attitudes are fully transparent to us. One of the virtues of this view is that it can explain both why there is pressure to re-interpret putatively incoherent agents (as the tendency to be coherent is built into what beliefs, intentions and other relevant attitudes are) and why it is nevertheless possible to be incoherent (as dispositions can be defeated and conditions of full transparency are not always met).

Chapter 6 argues that the different types of structural irrationality correspond with *requirements* of structural rationality, which are 'wide-scope-in-spirit' because they are prohibitions on attitude combinations that do not privilege particular ways of resolving an incoherence. Chapter 7 shows that this view is not committed to an often criticized wide-scope semantics of ordinary conditionals that are regularly associated with structural requirements. In Chapter 8, Worsnip completes his theory of structural rationality by defending a novel view of its normative significance. In line with some other authors, he maintains that facts about what structural rationality requires (or, alternatively, facts about incoherence that ground these facts) provide reasons. Yet three distinctive features of Worsnip's account are supposed to render it immune against the much-discussed criticisms that Kolodny and others have put forward against such views. The first is that the reasons in question are *non-derivative*. This avoids the difficulty, shared by all broadly instrumentalist approaches, of having to show that being structurally rational is, in each particular instance, conducive to some other objective that we have reason to pursue, such as autonomy or prudence. The second feature is that the reasons are only *indirectly related* to the requirements: they aren't reasons for the (sets of)

attitudes that structural rationality requires, but rather reasons to *deliberate* in certain ways. In Worsnip's words:

> *Reasons-to-Structure-Deliberation Model.* Considerations of coherence constitute reasons to structure deliberation in certain ways. More specifically: the fact that some possible combination of attitudes is incoherent is a reason to treat it as off-limits in one's deliberation. (256)

As a result of this second feature, the account promises to avoid all the problems stemming from the commonly shared assumption that reasons of structural rationality would have to be reasons for *attitudes*. In particular, it promises to avoid the problem that they would be state-given reasons for attitudes that seem to be alien to the way we typically deliberate, if it is even possible to deliberate with them at all. Finally, the third feature is that the reasons are fittingness-related rather than value-related. This means that Worsnip can avoid potentially implausible implications to the effect that structuring one's deliberation in the suggested way has final value. It is also supposed to show that the reasons in question are 'right-kind' rather than 'wrong-kind' reasons – a move that is made possible by the second feature, as the reasons *would* be wrong-kind *if* they were reasons for attitudes.

In the third and last part of the book (which consists in a final chapter and a brief coda), Worsnip helpfully connects the issue of structural rationality to a number of other ongoing philosophical debates (including, but not limited to, disputes on moral rationalism, higher-order evidence, epistemic permissivism, and the relation between normativity and value) and shows how his arguments have substantial implications for them.

In the remainder of this review, I will discuss some parts of Worsnip's theory in a bit more detail and raise some questions about it. Although other aspects of the book would be equally worth engaging with, I will here focus on Worsnip's defence of dualism (2), and the question of normative significance (3).

2. Dualism vs. reasons-based monism

Dualism comprises three claims about structural and substantive rationality: that they are distinct, that they are 'genuine kinds of rationality' (4), and that they cannot be reduced to each other. My first worry concerns the *genuineness thesis.* On a natural reading, one would expect

this thesis to entail 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. But while Worsnip dedicates a lot of space to showing that *structural* rationality is unified, the question of what unites structural and substantive rationality as *kinds of rationality* is not addressed in his book, and nothing suggests that Worsnip takes himself to be committed to providing an answer to it.

But what does the genuineness thesis amount to if there is nothing that holds structural and substantive rationality together? It is difficult to see how it could be more than the expression of a terminological choice to use the word 'rationality' in two different senses. And while that decision may itself be unobjectionable, the terminological reading of the genuineness thesis reveals that Worsnip's dualism does not differ in substance from the coherence-based position of Broome and others, which reserve the term 'rationality' for the structural phenomenon and are therefore classified as 'monist' by Worsnip (cf. 99). As a result, the classification itself becomes dubious, at least insofar as we have reason to categorize views along the lines of important rather than unimportant distinctions. As far as I can see, the relevant distinction is the one between (i) a dualist view that takes structural rationality and reasons to be independent of each other, and (ii) a monist view that understands one of these notions in terms of the other. Whether those who hold the first of these views should use the term 'rationality' to refer to both of the phenomena that they take to be independent of each other, as Worsnip holds, or to only one of them, as Broome maintains, does not strike me as an important question.

A second worry I had was that Worsnip is not always as careful as he should be in distinguishing the *distinctness thesis* from *dualism* – a view that entails the *irreducibility thesis* in addition to the distinctness thesis. The distinction is important because, as Worsnip notes himself (58), reductionists who appeal to the Guarantee Hypothesis can account for the distinctness of structural rationality; they can understand structural irrationality as a failure to respond correctly to reasons that is distinct because it is *guaranteed by combinations of attitudes* (cf. Kiesewetter 2017: 238–9). And yet the book contains several passages in which Worsnip purports to support dualism by appealing to considerations that really only support the distinctness thesis (cf. 4–6; 281–8; 302). And there are other passages in which Worsnip downplays philosophical worries about the dualist picture (which are discredited as 'anxiety attacks') by appealing to the harmlessness of the distinctness thesis (compare: 'all that dualism,

in itself, commits us to is the claim that it's possible to have gone wrong in one way [viz., substantively] without having gone wrong another way [viz., structurally]', 24).

This brings me to the third worry, which concerns the dialectical situation between Worsnip's dualism and the reasons-based view. As far as I can see, all the initial attractions of dualism that Worsnip brings forward turn out to be attractions of *any* view that can account for the distinctness of structural rationality, and therefore do not favour dualism over versions of the reasons-based view that accommodate the distinctness thesis. On the other hand, it seems to me that a monist reasons-based view has a number of attractions that dualism cannot claim for itself, even putting the obvious advantage of parsimony aside. For example, it accounts not only for the unity of structural rationality, but also for the unity of structural and substantive rationality, and it also provides a plausible explanation of why ascriptions of irrationality amount to criticism.

These general considerations do not defuse the challenges that Worsnip raises for the Guarantee Hypothesis, but I think they show them in a slightly different light. The method of reflective equilibrium, which Worsnip commits himself to in the context of his own theory of structural rationality (128–9), aims to preserve core intuitions about paradigmatic cases, but it also allows for the revision of less central and reflectively less stable pre-theoretical judgements or intuitions in light of attractive theoretical assumptions that lead to a more unified and explanatorily powerful theory. My impression is that Worsnip's rejection of the Guarantee Hypothesis is not based on core intuitions about paradigmatic cases and that it generally does not take into account the theoretical attractions of the reasons-based view.

I lack the space to discuss this in detail, but let me at least illustrate this point by one example. Worsnip argues that the Guarantee Hypothesis fails in cases of cyclical preferences that are individually permitted by our reasons (for example in cases of value incommensurability). But how much weight should we attach to the intuition that cyclical preferences are irrational in such cases? It's not clear to me that we are dealing with a core intuition about a paradigmatic case that must be considered indispensable in reaching reflective equilibrium. In typical cases, preferences go along with betterness-judgements, and cyclical betterness-judgements arguably cannot be individually supported by reasons. Moreover, the money pump argument shows that there are strong prudential reasons against making choices in accordance with cyclical preferences. How much weight should we attach to the judgement that cyclical preferences are irrational even in cases in which they are not based

on betterness-judgements and in which they do not lead to cyclical intentions? Even if it appears initially plausible, it seems to me clearly the kind of pre-theoretical judgement that we might well revise on reflection if it conflicts with an otherwise attractive, unified and explanatorily powerful theory.

Worsnip (88-90) raises four problems for this 'exception-making strategy', as he calls it, but none of them strike me as very serious. The first is that allowing for cases of rational incoherence creates the need for an account of coherence that is independent of the account of structural rationality. This is true, but it's not clear why it should be considered a problem or objection. Moreover, if Worsnip's account of incoherence is correct, proponents of the reasons-based view can simply adopt it. The second problem is that the proposal fails to show that '*all* the patterns of attitudes associated with structural irrationality guarantee substantive irrationality' (89, original emphasis). This overlooks both that the question of which patterns of attitudes are irrational is often debatable and that the method of reflective equilibrium allows for revising less central and less stable pre-theoretical judgements. Worsnip's third and fourth problems are based on the worry that allowing for exceptions is inconsistent with the assumption that structurally irrational attitudes can be detected on the basis of their general form rather than their particular content. However, as long as the exceptions (e. g., about cyclical preferences) are explained in terms of other, exceptionless and purely structural conditions (e. g., about cyclical betterness-judgements or intentions), that assumption is preserved.

The point about reflective equilibrium can be reinforced by considering the fact that, when it comes to his own theory of structural rationality, Worsnip is explicitly willing to make considerable sacrifices 'in order to achieve a well-unified theoretical account' (129). For example, since his view does not allow for rational incoherence, it rules out the intuitive phenomenon of a 'rational delay' (Podgorski 2017) in adjusting incoherent attitudes (187). And even though he 'prefer[s] to avoid it if possible', Worsnip seems willing to accept the revisionary conclusion that akrasia is not structurally irrational (144). (The problem with akrasia is that on Worsnip's account, for akrasia to be irrational, it needs to be part of the nature of ought-judgements that they involve dispositions to intentions – an assumption that seems to conflict with cognitivism about normative judgements.) Note that akrasia is often considered to be a paradigmatic case of structural irrationality (Scanlon 2007; Broome 2013), and that the Guarantee Hypothesis provides an elegant explanation of this assumption (Kiesewetter 2017,

246–48). In sum: It seems to me far from clear that Worsnip's own theory has a better balance of costs and benefits than the reasons-based view. Notably, this is true even if we focus on the theories' ability to account for structural rationality *alone* and bracket the independent advantages of the reasons-based view as a unified and normative theory of rationality *simpliciter*.

3. The normative significance of structural rationality

Let me now turn to Worsnip's account of the normative significance of structural rationality. I'll start with a worry about Worsnip's notion of a requirement of structural rationality, before turning to his proposal that such requirements constitute non-derivative reasons for structuring one's deliberation.

Proponents of structural requirements of rationality who consider it to be a substantive normative question whether we have reason to comply with such requirements face the challenge of explaining what they mean by saying that structural rationality *requires* something (Way 2010, 1065; Kiesewetter 2017, 42–44). They cannot mean that there are (decisive) reasons to do it. But they also typically accept that they don't mean to use 'requirement' in the sense of a conventional code or necessary condition. Worsnip seems to accept this challenge and sees his account of normative significance as meeting it (30–31). But there are at least two reasons to be skeptical that the account is in the position to do that. The first is due to the indirectness of Worsnip's account, which entails that the requirements of structural rationality and their corresponding reasons have different contents – (absences of) attitude-sets on the one hand, and reasoning activities on the other. Now, it is relatively clear how an appeal to reasons for F-ing might help to illuminate the sense in which F-ing is required. But it is far less clear how reasons to do something *other* than F-ing helps to illuminate the sense in which F-ing is required.

The second reason to be skeptical is that Worsnip takes the relevant reasons to be facts about structural requirements. This seems to imply that an understanding of the reasons of structural rationality requires an independent understanding of requirements of structural rationality, one that doesn't appeal to the normative significance of structural rationality. The question is: What are those facts that according to Worsnip constitute reasons to structure one's deliberation? One cannot answer that question by pointing out that the facts provide reasons to structure one's deliberation.

Setting this issue to one side, let us consider how plausible Worsnip's suggestion is that we have non-derivative reasons to treat incoherent attitude-combinations as off-limits in deliberation. Why should we accept this thesis? Consider an analogous view about the normative significance of grammar, according to which we have non-derivative reasons against using ungrammatical linguistic expressions. I hope we can all agree that this is not a very plausible view. Surely, the reasons we have to express ourselves in ways that others understand will very often give us reasons for heeding grammatical rules. But such reasons aren't always present, nor do they necessarily speak against every possible violation of grammar. The proponent of the normative significance of grammar might reply: 'You misunderstood me. I did not mean to say that there are necessarily *derivative* reasons for complying with the rules of grammar. My view is that facts about such rules constitute *non-derivative* reasons for speaking correctly.' What should we say about this reply? As Worsnip points out, everyone who accepts any reason will, on pain of regress, at some point have to accept that some fact F constitutes a non-derivative reason to G, which means that in such cases the question 'Why G?' will not receive a more informative answer than 'Because F' (cf. 266). But obviously, this cannot mean that all assumptions about non-derivative reasons are equally warranted. We need good reasons to accept such an assumption, especially if it isn't intuitively obvious and if it is disputed in the dialectical context. What better reason do we have to accept the view that structural requirements constitute non-derivative reasons than we have to accept the view that the rules of grammar constitute non-derivative reasons?

The answer cannot lie in any presumed intrinsic value that is carried by treating incoherent attitude combinations as off-limits but not carried by avoiding ungrammatical linguistic expressions. Worsnip concedes that it is not plausible to think that non-derivative reasons of structural rationality could be value-based but he also rejects the view that all reasons must be value-based (I happen to agree, see Kiesewetter 2022). Moreover, Worsnip holds that value-based reasons are to be identified with wrong-kind reasons, and his stated aim is to defend the view that structural requirements constitute *right-kind* reasons, which he takes to be 'fit-related', i.e., constitutively connected to the fittingness of the favoured response (cf. 260).

Worsnip's argument that structural requirements constitute right-kind reasons is that 'it is fitting to structure deliberation in ways that respect coherence constraints' (261). To support this, he offers two alternative understandings of fittingness, one in terms of correctness relative

to constitutive standards, and another one in terms of responses to objects that merit or are worthy of them. He then claims, firstly, that 'it is constitutive of deliberation that it's correct to treat incoherent combinations of attitudes … as off-limits', and secondly, 'that an incoherent combination of attitudes *merits* being treated as off-limits'. Suppose we accept all that. This might show that a presumed reason to structure deliberation in the suggested way would be a right-kind rather than a wrong-kind reason, but it does not show that such a reason exists in the first place. To reach this conclusion, we have to assume that there is a normative reason to perform an action if that action is fitting. But that does not seem right, at least not if we understand fittingness in terms of correctness standards. Suppose it's constitutive of a certain game that it's correct to make a particular move. There may still be no reason to make the move. If you can save someone's life by making the incorrect move, there is no 'fit-related' normative reason that competes with the reason for making the incorrect move (cf. Thomson 2008, 90). Even in less dramatic circumstances, you may have no reason to make the correct move, unless you have a reason to play the game in the first place (some authors, such as Lord and Sylvan (2019), believe that there are *non-normative* reasons to comply with constitutive standards, but such reasons could obviously not help to vindicate the normative significance of structural rationality). Similar points apply to correctness standards for the use of language.

Things might look better if we do not understand fittingness in terms of correctness (cf. Howard and Leary 2022). Plausibly, there is a notion of 'fitting action' that entails a reason to act, and according to which it wouldn't be fitting to make the correct move if one doesn't have a reason to play the game. What is difficult to see, however, is how one could claim that structuring one's deliberation in certain ways is fitting in this sense without *presupposing* that there is reason to do so.

Thus, either way we are left without an argument for Worsnip's claim that there is the kind of non-derivative fit-related reason on which he builds the normative significance of rationality. If we understand fittingness in terms of correctness, we should deny that there are any non-derivative fit-related reasons for action, because correctness standards quite generally do not generate non-derivative reasons for action. If we understand fittingness independently of correctness, we cannot assume that it's fitting to structure one's deliberation in the suggested ways without presupposing what is at issue, namely that there are non-derivative reasons to do so. As far as I can see, then, we are in no better position to believe that there are non-derivative reasons to treat incoherent combinations of attitudes as off-limits in

deliberation than we are to believe that there are non-derivative reasons to treat ungrammatical expressions as off-limits in conversation. In my view, this casts serious doubt on Worsnip's account of the normative significance of structural rationality.

4. Conclusion

I found myself in disagreement with many of the views and arguments that Worsnip presents, and for obvious reasons I have focused on these disputes in this review, but this should not be taken to reflect badly on the quality of this extraordinary book. *Fitting Things Together* provides the most convincing account of structural rationality as an independent and normatively significant phenomenon to date. It presents an illuminating characterization of structural rationality and a compelling case for its distinctness, an original and sophisticated theory of structural rationality as a unified phenomenon, and a novel and intriguing account of its relevance. These are significant philosophical achievements that will serve as important reference points for future work on rationality. Over and above all that, *Fitting Things Together* is a book written with masterly skill, admirably clear and with a great sensitivity for weighing the need for details against the benefits of focusing on the bigger picture, which makes it highly accessible even for those who haven't yet dealt with the somewhat detail-obsessed literature on the subject. It is compulsory reading for philosophers working on rationality, but also highly recommended for anyone who wants to get a grip on what is at issue in the debate.[*]

BENJAMIN KIESEWETTER
Bielefeld University, Germany

[*] I am very grateful to John Broome, Zoë Johnson King, Felix Koch, and Alex Worsnip for valuable feedback on an earlier draft of this review. Work on this review was funded by the European Union (ERC Grant 101040439, REASONS F1RST). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Broome, John. 2013. *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.

Howard, Christopher, and Stephanie Leary. 2022. 'In Defence of the Right Kind of Reason'. In *Fittingness*, edited by Chris Howard and R. A. Rowland, 221–42. Oxford University Press.

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

———. 2022. 'Are All Practical Reasons Based on Value?' *Oxford Studies in Metaethics* 17: 27–53.

Kolodny, Niko. 2005. 'Why Be Rational?' *Mind* 114 (455): 509–63.

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

Lord, Errol, and Kurt Sylvan. 2019. 'Reasons: Wrong, Right, Normative, Fundamental'. *Journal of Ethics and Social Philosophy* 15 (1): 43–74.

Podgorski, Abelard. 2017. 'Rational Delay'. *Philosopher's Imprint* 17 (5): 1–19.

Scanlon, T.M. 2007. 'Structural Irrationality'. In *Common Minds. Themes from the Philosophy of Philip Pettit*, edited by Geoffrey Brennan, Robert Goodin, Frank Jackson, and Michael Smith, 84–103. Oxford: Oxford University Press.

Thomson, Judith Jarvis. 2008. *Normativity*. Chicago and La Salle: Open Court.

Way, Jonathan. 2010. 'The Normativity of Rationality'. *Philosophy Compass* 5 (12): 1057–68.