

When Things Fail to Fit Together

Daniel Fogal

Penultimate version - please cite published version where possible

1. (Not) Fitting Things Together

Much of our mental life is a mess. We believe various things at various times and other things at other times, sometimes as the result of thoughtful deliberation but oftentimes not. Our beliefs vary widely in their significance, strength, basis, and content—we have random, fleeting perceptual beliefs, significant but slowly eroding memorial beliefs, foundational and firmly held religious beliefs, and everything in between. And what goes for belief goes for many of our other attitudinal mental states (or ‘attitudes’, for short): our intentions, preferences, hopes, fears, and the like. It’s no wonder, then, that such attitudes often fail to fit together, and rarely if ever coalesce into a stable, unified whole.

Our attitudes can fail to fit together in two main ways. One is for them to *conflict* or otherwise be in tension with one another. Sometimes we end up with inconsistent beliefs, for example, or with beliefs that go against our assessment of the evidence. Other times we end up with intentions that can’t be jointly satisfied, with desires that we think we shouldn’t have, or with fears that we take to be misplaced. The list of possibilities goes on. In cases like this our attitudes don’t just fail to fit together, they positively clash, and the tension is only resolved when one or more of the offending attitudes is revised. In the recent philosophical literature, cases of inconsistent beliefs, inconsistent intentions, akrasia, cyclical preferences, and the like have been treated as paradigmatic examples of *incoherence* and (what’s been called) *structural irrationality*.

The second, more common way for our attitudes to fail to fit together is for them to simply be *unrelated* to one another—to be a random assortment of attitudes of various kinds. I believe that $2 + 2 = 4$, that I’m a father, and that tigers have stripes. I intend to visit Sweden again, buy groceries this weekend, and eventually write a book. I fear that climate change is getting worse, hope that the Lakers will win the NBA championship, and prefer vanilla over chocolate. None of these attitudes are connected or otherwise meaningfully related to each other, aside from the fact that they belong to the same person—namely, me. They don’t clash, but they also don’t fit together or (as I’ll usually put it) *cohere*.

What exactly it means for our attitudes to clash or cohere is a difficult question. But what matters for now is simply the observation that attitudinal coherence—i.e., the fitting-togetherness of attitudes—requires more than merely the absence of incoherence. Random collections of attitudes that fail to clash do not thereby cohere.

Why does this seemingly obvious observation matter? Because much of the debate over the nature and significance of coherence and structural rationality over the past twenty years overlooks it—it’s a debate that focuses almost exclusively on incoherence, and coherence (insofar as it is discussed at all) is treated as the absence of incoherence. The intuitive reflections above, however, cast significant doubt on this assumption.

The main purpose of this essay is not to critically engage with the expansive literature on attitudinal coherence and structural rationality in detail. (I do that elsewhere.¹) Instead, it’s to illustrate the significance of the observation that coherence is more than merely the absence of incoherence by pursuing an extended case study, focusing specifically on Alex Worsnip’s *Fitting Things Together*.²

To set the stage: one of the most prominent trends in normative theorizing has been the increasing focus on reasons, rationality, and the relationship between them. This trend has resulted in a steady stream of articles as well as books, with one of the main fault lines being whether rationality is a unified affair—those who answer affirmatively are *monists* about rationality—and if so what it consists in. Some monists take rationality to solely be a matter of attitudinal coherence, and hence equate structural rationality with rationality *simpliciter* (cf. Broome 2013), while others take rationality to fundamentally be a matter of responding correctly to one’s normative reasons, and hence equate rationality with (what is called) *substantive* rationality (cf. Kiesewetter 2017, Lord 2018).³

Worsnip, by contrast, offers a forceful defense of *dualism* about rationality—the view that structural and substantive rationality are “two distinct but equally genuine kinds of rationality, neither of which is reducible to the other” (2021: 4). The initial motivation for dualism appeals to intuitions about cases (cf. Pryor 2018; Way 2018; Fogal 2018). Consider Tom, for example, who believes that he’s Superman and that Superman can fly but doesn’t believe that he himself can fly—indeed, suppose he’s tried and failed multiple times, and so is convinced he can’t.⁴ Tom is clearly irrational, and in two seemingly distinct ways: he’s *substantively* irrational in virtue of believing he’s Superman (which, we may safely suppose, isn’t supported by his evidence) and *structurally* irrational in virtue of having inconsistent beliefs. The second rational failing, unlike the first, doesn’t seem to have anything to do with Tom’s evidence or reasons—if anything, Tom is *correctly* responding to his evidence in believing that he himself can’t fly. Instead, the failing seems distinctively structural in nature, having to do with the (im)proper relationships between Tom’s *attitudes themselves*, rather than between his attitudes and reasons.

¹ See, for example, Fogal 2020, Fogal & Worsnip 2021, and Fogal & Risberg ms.

² Worsnip, A. 2021. *Fitting Things Together. Coherence and the Demands of Structural Rationality*. New York: Oxford University Press.

³ Other notable book-length discussions of rationality and its (non-)relationship to reasons include Wedgwood 2017 and Brunero 2020.

⁴ This example originated with Jim Pryor.

To strengthen the intuition, consider Tom's older brother Tim, who like Tom believes that he's Superman and that Superman can fly, but unlike Tom makes the obvious inference and comes to believe that he himself can fly. As Worsnip observes:

"There's a clear sense in which Tim is even less rational than Tom, since he has two beliefs that go dramatically against his evidence (viz., that he is Superman, and that he can fly), where Tom has only one (viz., that he is Superman). But there is also a clear sense in which Tim is more rational than Tom, since his beliefs cohere in a way that Tom's don't: there's no inconsistency in his beliefs... [W]e can recognize both of these senses by saying that Tom is less substantively irrational than Tim, but Tim is less structurally irrational than Tom." (2021: 6)

Analogous cases (and pairs of cases) arise in the practical realm involving intentions, preferences, and the like. Hence the initial case for dualism about rationality, both theoretical and practical.

After sketching the initial case for dualism, Worsnip proceeds to defend it at length in Part I of the book (Chapters 1-4). Part II (Chapters 5-8), in contrast, focuses squarely on the structural side of things, with Worsnip providing novel accounts of the nature of incoherence and the normativity of structural (ir)rationality, as well as of our ordinary thought and talk about such. Part III (Chapter 9, Coda) then explores a variety of "upshots" for other debates, including debates concerning moral rationalism, rational choice theory, higher-order evidence, the normativity of logic, epistemic permissivism, and Bayesianism, with a concluding coda encouraging us to resist the "tyranny" of value.

As this brief summary makes clear, *Fitting Things Together* is ambitious in its aims and wide-ranging in its content. It is also admirably nuanced throughout—it brims with important distinctions, helpful clarifications, and careful arguments, and does so without being bogged down by them. Indeed, it's not much of an exaggeration to say that, in many ways, Worsnip's book represents contemporary analytic philosophy at its best. So there's a lot—or, more accurately, *a lot*—to like about *Fitting Things Together*.

There's just one problem: contrary to what one might expect, *Fitting Things Together* is *not* about fitting things together. It's about one of the ways in which one might *fail* to fit things together—namely, by having attitudes that clash with each other. So it's a book about *incoherence*, not coherence. This is evident from the outset—the initial list of illustrative examples all involve incoherence (2021: 3)—and reinforced by the index, where the entry for 'coherence' simply says "see *incoherence*" while the entry for 'incoherence' is loaded with sub-entries and page numbers. As suggested above, however, coherence—i.e., the fitting-togetherness of attitudes—is a positive notion that requires more than merely the absence of incoherence (in the sense at issue), much like being good requires more than merely not being bad and being happy requires more than merely not being sad. In each case there's a neutral, non-valenced state—i.e. being neither coherent nor incoherent, being neither good nor bad, being neither happy nor sad—between the positively- and

negatively-valenced ones. As a result it would be a mistake to equate either of the latter with the absence of its oppositely-valenced state—to do so would collapse the difference between being valenced (e.g. being happy, being sad) and being neither-positively-nor-negatively-valenced (being neither happy nor sad).

We can make the concern more concrete by returning to the case of Tom and Tim, and introducing a new character: their sister Tam. Like both of her brothers, Tam believes Superman can fly, but unlike them she doesn't believe that she's Superman. In fact, one of the only other beliefs she has about Superman is that he loves pasta; she otherwise prides herself on her ignorance of all-things-Superman. Compare the following sets of beliefs:

Incoherent Tom:	{I'm Superman, Superman can fly, I can't fly}
Coherent Tim:	{I'm Superman, Superman can fly, I can fly}
Random Tam:	{I don't care about Superman, Superman can fly, Superman loves pasta}

Intuitively, the three combinations of beliefs differ not just in content but also in coherence (considered on their own, in isolation from other possible beliefs the subjects may have). Whereas Tom's beliefs (unlike Tim's and Tam's) *clash* or *conflict*, Tim's beliefs (unlike Tom's and Tam's) *positively cohere* or *fit well together*. And although Tam's beliefs all concern Superman, they are otherwise (unlike Tom's and Tim's) *random* or *unrelated*.

However, on Worsnip's view, since a set of beliefs is coherent just in case it's not incoherent, there's no difference, structurally speaking, between Tam and Tim—neither set of beliefs is incoherent, and so both count as coherent. This strikes me as inadequate to the phenomenon. Worsnip's view of coherence is therefore either false or incomplete. It's false if it's understood as providing an account of the fitting-togetherness-of-attitudes—or what we can call *strong* coherence. It's incomplete if it doesn't, and instead merely concerns the absence of clashing attitudes—or what we can call *weak* coherence.⁵ Either way, while Worsnip's view can say what's wrong, structurally speaking, with Tom, it's insensitive to the difference between Tim and Tam. It thus fails to do justice to the full range of our intuitions about the relevant cases, and in particular to our intuitions about strong coherence, the oppositely-valenced counterpart of incoherence.

So that's the bad news: *Fitting Things Together* is not about fitting things together. But there's also good news: Worsnip's novel account of incoherence can be naturally extended to provide an account of strong coherence (henceforth just 'coherence'). I'll elaborate and defend this claim in §2. This then leads to more bad news: the account of coherence is at odds with Worsnip's preferred, requirements-centric account of structural rationality. But, as I'll argue in §3, this is not bad news *tout court*, since we should reject the requirements-centric picture anyway and opt for a "pressure-based" account in its stead.

⁵ In personal communication, Worsnip has indicated that this is his preferred response. For a fuller response to his response, see Fogal and Risberg ms.

I'll then conclude in §4 by considering and replying to one of Worsnip's main objections to the pressure-based account.

2. From Incoherence to Coherence

In Part I (Chapters 1-4) Worsnip says a lot about what structural rationality isn't and not—by his own admission—as much about what it *is*. The positive account is instead provided in Part II (Chapters 5-8), where Worsnip offers a novel account of incoherence and structural irrationality, which he takes to be co-extensive properties with the former grounding the latter—structurally irrational combinations of attitudes are irrational *in virtue of* being incoherent. He begins, in Chapter 5, by giving a general account of what unifies instances of incoherence, and hence structural irrationality. The motivation for such an account is clear: given that the instances of incoherence are so diverse, we should try to find a unified account of them if we can. In particular, Worsnip employs the method of “reflective equilibrium” in looking for a property shared by all and only incoherent states that can be used as a kind of test for incoherence (2021: 128). He is officially neutral, however, on whether the property tells us what incoherence *is*, and hence amounts to an analysis.

Worsnip's account starts with an important but often-overlooked observation: incoherent attitudes are *puzzling*. After asking the reader to consider cases of clear-eyed means-end incoherence, explicit commitment to incompatible courses of action, conscious endorsement of cyclic preferences, actively fearing *A* while being convinced that *A* is good, and so on, Worsnip notes:

“[I]t's hard to make sense of agents who stably sustain incoherent states while being fully and reflectively aware of those states... Instead, we tend to reach for descriptions of putatively incoherent agents either (i) as genuinely having the incoherent attitudes, but in some way suffering from some failure of transparency—where this includes failures of self-knowledge, instances of self-deception, fragmentation, attitudes being buried ‘beneath the surface’ or held merely implicitly, and the like—or (ii) as not in fact having the incoherent attitudes at all, but in fact being in ‘weaker’ states (such as desire rather than intention, or hope rather than belief).” (2021: 132)

Worsnip then uses this observation as the basis of the following incoherence “test”:⁶

⁶ I'm following Worsnip in formulating the relevant test as a biconditional, but another (perhaps ultimately preferable) possibility is to treat it as a useful heuristics or generally-but-not-perfectly reliable rules of thumb. The latter, unlike the former, is not threatened by some of the more nitpicky counterexamples. The biconditional formulations matter most for those seeking to provide an analysis or reductive definition, as opposed to those with more modest philosophical ambitions.

Incoherence Test

A set of attitudinal mental states is jointly incoherent if and only if 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)

By ‘conditions of full transparency’ Worsnip means “conditions under which the agent knows, and explicitly and consciously believes, that she has the states in question, without self-deception, fragmentation, or any failure of self-knowledge (pertaining to those states), and simultaneously with respect to each state in the set” (2021: 133).⁷ And by ‘disposition’ Worsnip means a *strong but (in principle) defeasible* disposition, since “even relatively strong dispositions can be defeated” (2021: 142, fn 16).

The Incoherence Test classifies many paradigmatic instances of incoherence as incoherent. Suppose, for instance, you intend to go on a run, believe that going on a run requires getting out of bed, but do not intend to get out of bed. Suppose further that you’re aware that you have all three attitudinal mental states. (I’ll be following Worsnip in using ‘attitudinal mental states’ in a broad sense that includes the absence of attitudes.) Question: are you guaranteed to be disposed to revise at least one of the (first-order) states? Plausibly, yes. And is your disposition to revise constitutive of (at least some of) the initial states in question? Again, plausibly yes—were one to *not* be disposed, even under conditions of full transparency, to either (a) drop the intention to run, (b) form the intention to get out of bed, or (c) revise the belief that running requires getting out of bed, then it’s hard to see how you could count as genuinely intending to go for a run while believing that doing so requires getting out of bed. More generally, it seems constitutive of intention that one be disposed to revise the intention to *A* or else form the intention to *B* when one believes that *A*-ing requires *B*-ing, at least when one is fully aware of having such attitudes. Something similar seems true of many other cases of incoherence and structural irrationality, such as:

Belief-contradictoriness:	{believing that <i>p</i> , believing that not- <i>p</i> }
Intention-contradictoriness:	{intending to <i>A</i> , intending to not <i>A</i> }
Cyclic preferences:	{preferring <i>A</i> to <i>B</i> , preferring <i>B</i> to <i>C</i> , preferring <i>C</i> to <i>A</i> }
Inter-level incoherence:	{believing that one’s evidence decisively supports <i>p</i> , not believing that <i>p</i> }

⁷ As Berker ms notes, the requirement that you *know* that you have the relevant attitudes is arguably unnecessary—all that’s needed is the assumption that you *occurrently* (i.e. explicitly and consciously) *believe* that you have the relevant attitudes. Switching from conditions of full transparency to conditions of “simultaneous occurrency”—i.e., conditions under which all of the relevant attitudes are, at the same time, occurrent (rather than merely dispositional)—also helps avoid various counterexamples to Incoherence Test, including ones in which some of your attitudinal mental states make one of your beliefs false, such as {believing that one intends to *A*, not intending to *A*}.

Despite its success in classifying cases like that above, there are a variety of concerns about the Incoherence Test one might have. Worsnip considers many of the most obvious and/or pressing ones, and does an admirable job addressing them. But what about coherence? Might we be able to construct a comparably plausible Coherence Test? Perhaps so. One simple suggestion, for example, would be to replace the appeal to *revision* with an appeal to *retention*. Modifying the Incoherence Test, we get:

Coherence Test A set of attitudinal mental states is jointly coherent if and only if 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 retain the full set of states.

Though useful as an illustration of how Worsnip's view might be extended, there are reasons to doubt the Coherence Test. Return to Coherent Tim, for example. It seems doubtful that it's constitutive of Tim's beliefs that, when conditions of simultaneous occurrency are met, he is disposed to retain the full set of beliefs. Instead, it seems that Tim's belief that he's Superman together with his belief that Superman can fly disposes him (constitutively) to retain his belief that he can fly, insofar as he retains his first two beliefs. But that doesn't amount to having a disposition to retain all three beliefs *as a whole*.⁸ This verdict points away from the Coherence Test and towards my preferred alternative, the Support Test, discussed in §3 below.

What matters most for present purposes, however, is that the combination of (something like) the Coherence Test and the Incoherence Test allows for attitudinal mental states to be neither coherent nor incoherent, such as sets of random, unrelated attitudes. When random, unrelated attitudes are fully transparent, for example, the relevant agent won't typically have a disposition to retain such sets of attitudes—at least not over and above whatever dispositions to retain each attitude come with the individual attitudes themselves—nor a disposition to revise them, and even if there is it won't plausibly be constitutive of any of the attitudes involved.

Importantly, both of the tests above can be naturally modified to allow for degrees of (in)coherence by appealing to the strength of dispositions to revise (in the case of incoherence) or retain (in the case of coherence). Worsnip recognizes the possibility of modifying the Incoherence Test to allow for degrees, and appeals to it in responding to an important objection—namely, that the test fits poorly with practical akrasia, since clear-eyed, transparent akrasia seems both widespread and at least far *less* puzzling than many of the other kinds of paradigmatic incoherence. Akrasia is incoherent, Worsnip suggests, but not *as* incoherent as, say, means-end incoherence.

⁸ Thanks to Selim Berker for raising this worry.

degrees, one might therefore expect structural irrationality to as well. But that's not what Worsnip says. Instead, he says that we "need to specify a threshold of incoherence required to be structurally irrational" (2021: 187, n. 44). Unlike incoherence, then, structural irrationality is treated as a non-graded, all-or-nothing affair. He further expresses sympathy with the view that the threshold of incoherence required for structural irrationality is "*any incoherence at all*" (ibid, italics in original). On Worsnip's view, then, any combination of attitudes that is incoherent to any degree at all will count as structurally irrational *full stop*.

This view isn't forced on Worsnip, though. Another option would be to say that both incoherence and structural irrationality come in degrees, with the degree of the former determining the degree of the latter. On this alternative, while any degree of incoherence suffices for there being *some* structural irrationality, the degree of structural irrationality (and the severity of violating the corresponding requirement) increases as the degree of incoherence increases—just as one might expect given that incoherence and structural irrationality are coextensive, with the former grounding the latter.¹¹

However, even if the wide-scope, requirements-centric account of structural irrationality can be reconciled with the graded nature of incoherence, more remains to be said about the positive side of structural rationality, and in particular its relationship to (strong) coherence. Indeed, given the close connection between structural irrationality and incoherence, it would seem natural to expect there to be a similarly close connection between structural rationality and coherence—or at least a story about why such an expectation is misguided. At a minimum, then, the wide-scope, requirements-centric account view requires elaboration and supplementation in order to fully account for our judgments concerning structural (ir)rationality.

In this regard (among others), Worsnip's account stands in stark contrast with the "pressure"-based account of structural rationality that I've developed elsewhere (Fogal 2020) and that Worsnip discusses in Chapter 6. The pressure-based view has no difficulty recognizing both the positive and negative sides of structural rationality, and it readily makes sense of our judgments concerning coherence, incoherence, and the absence of both. In the next section (§3) I'll briefly sketch the pressure-based view and construct Worsnip-style (in)coherence "tests" for it. I'll then conclude in §4 by replying to one of Worsnip's main objections.

3. From Requirements to Pressure

So far I've been following Worsnip and much of the rest of the literature on structural rationality in focusing on the *monadic* properties of coherence and incoherence as

view and his own, leaving open the (im)possibility of reconciling the claim that structural irrationality is grounded in the violation of requirements with the claim that structural irrationality is grounded in incoherence (2021: 169, n. 7).

¹¹ Thanks to Alex Worsnip for suggesting this alternative.

instantiated by *sets of attitudes* (or attitudinal mental states). But there are other (in)coherence-like properties one might be interested in. For example, one might be interested in better understanding what it is for an attitude (or group of attitudes) to *cohere with* one another, or for an attitude (or group of attitudes) to *clash with* another. These are relational properties, rather than monadic ones. Both are plausibly symmetric: if A coheres/clashes with B, then B coheres/clashes with A. Alternatively—or in addition—one might be interested in directional, non-symmetric (in)coherence-like relations. Jim Pryor (2004), for instance, talks about how beliefs can “rationally commit” you to having certain other beliefs, where this is distinct from what one has justification or reason to believe. (Pryor is thus concerned with “structural” rather than “substantive” rationality, though he doesn’t use those terms.) He illustrates the phenomenon as follows:

Suppose you believe Johnny can fly. This belief rationally commits you to the belief that someone can fly. If you’re not justified in believing that Johnny can fly, though, you need not have any justification for the further belief. You may even have plenty of evidence and be fully justified in believing that no one can fly. But your belief that Johnny can fly still rationally commits you to the belief that someone can fly. (2004: 364)

Pryor then weakens and generalizes the notion of rational (i.e. structural) commitment, introducing the notions of *support* and *opposition*, both of which are defeasible and come in degrees. The (mere) belief that some hypothesis *H* is the best explanation of evidence *E*, for example, structurally supports (but doesn’t necessarily commit you to) believing *H*, while the belief that an expert claims that not-*p* structurally opposes (but doesn’t necessarily prohibit you from) believing that *p*. In both cases the support/opposition is strong yet defeasible, and can vary in strength depending on various factors (e.g. how good/reliable you believe the explanation or expert is).

Notably, Pryor’s proposal can be extended further, applying not just to doxastic states like belief but to attitudes more generally—to intentions, preferences, hopes, fears, and the rest. I pursue this generalization at length elsewhere (Fogal 2020), arguing that each dimension of rational evaluation corresponds to a distinct kind of pro tanto rational *pressure* or *force*—substantive rationality with *substantive* (or ‘justificatory’) pressure and structural rationality with *structural* (or ‘attitudinal’) pressure.¹² On this “pressure-based” view, believing that *p* and that if *p* then *q* generates substantial (pro tanto) structural pressure to believe that *q*, just as having good reasons to believe that *p* and good reasons to believe that if *p* then *q* generates substantial (pro tanto) substantive pressure to believe that *q*. Similarly, intending to φ and believing that φ -ing requires ψ -ing generates substantial structural pressure to intend to ψ , just as having good reasons to φ and good reasons to believe that φ -ing requires ψ -ing generates substantial substantive pressure to intend to ψ .

¹² I originally used the terms ‘justificatory’ and ‘attitudinal’ but for the sake of terminological simplicity and clarity I now prefer ‘substantive’ and ‘structural’.

The main difference between kinds of rational pressure lies in what generates them—i.e. their grounds or source. Whereas substantive pressure is generated by one’s reasons, structural pressure is generated by one’s (mere) attitudes. They can thus come apart in both directions: just as there can be substantive pressure without structural pressure, so there can be structural pressure without substantive pressure. The two kinds of rational pressure are otherwise very similar in nature. In particular, both:

- (a) are *force-like*, having something like magnitude (strength), which can vary, and direction (for/against);
- (b) enter into a variety of complex relations, and in particular relations of *support*, *opposition*, and *undermining*, all of which come in degrees; and
- (c) determine the rational status of one’s attitudes, and hence are *more fundamental* than “threshold-y” facts about (ir)rationality.

Although a lot could be said about each, the relationship between structural pressure and structural rationality—that is, (c)—is most relevant to our purposes. According to one natural account, being structurally rational is (roughly) a matter of doing—i.e. believing, intending, etc.—what the balance of structural pressure supports, just as being substantively rational is a matter of doing what the overall balance of substantive pressure supports—or, more colloquially, what you have most reason to do. In neither case is an attitude’s being rational a matter of an agent responding correctly to any individual amount or source of pressure considered on its own. Instead, in both cases what determines the rational status of attitudes are global, all-things-considered facts about rational pressure—facts which are in turn determined by interactions between various local, graded facts about (pro tanto) rational pressure. Thus, if the structural support for attitude A_1 generated by attitude(s) $A_{2,\dots,n}$ is (on balance) sufficiently strong, then A_1 will be structurally rational and, at the limit, required. Likewise, if the structural opposition to attitude A_1 that attitude(s) $A_{2,\dots,n}$ generate is (on balance) sufficiently strong, then A_1 will be structurally irrational and, at the limit, prohibited. Structural commitment can thus be plausibly understood as a special case of structural support—i.e. decisive, undefeated structural support—just as structural “prohibition” can be understood in terms of structural opposition—i.e. decisive, undefeated structural opposition.

What should we say about attitudes that are neither structurally supported nor structurally opposed by one’s other attitudes? One option is to say that such attitudes are structurally rational in virtue of not being opposed; another is to say that they’re structurally irrational in virtue of not being supported. Yet a third is to say they’re neither structurally rational nor irrational. My own preference, however, is to say that they’re structurally rational in a “weak” sense but not in a “strong” sense, where to be structurally rational in the weak sense is (roughly) to *not be structurally opposed* and to be structurally rational in the strong sense is (roughly) to be *structurally supported*. This is similar to the distinction between weak (or “negative”) and strong (or “positive”) permission commonly

drawn in deontic logic: in the weak sense, being permissible is just a matter of *not being prohibited*, while in the strong sense being permissible is a matter of being *positively sanctioned* (cf. von Wright 1963, among others).¹³

Let's turn now to the relationship between structural pressure and the properties of (in)coherence. On the pressure-based view, (in)coherence comes in degrees, with the degree of (in)coherence of a set of attitudes being a function of the support-relations and opposition-relations obtaining between the attitudes in the set. Since the function will likely be complex, there won't be a simple, straightforward way to determine the (in)coherence of large sets of attitudes. In general, though, we can say that a set of attitudes is pro tanto *coherent* to the extent that the attitudes in it either support or are supported by the others, and pro tanto *incoherent* to the extent that the attitudes in it either oppose or are opposed by the others. We can then define non-graded notions of (in)coherence in terms of the graded notion. For example, we can say that a set of attitudes is coherent *simpliciter* if it's pro tanto coherent to a sufficiently high degree and not pro tanto incoherent to any degree, and incoherent *simpliciter* if it's pro tanto incoherent to a sufficiently high degree and not pro tanto coherent to any degree.

We're now in a position to capture the intuitive verdicts concerning simple cases, including the ones typically at issue. Recall, for example:

Belief-inconsistency:	{believing that <i>p</i> , believing that not- <i>p</i> }
Intention-inconsistency:	{intending to <i>A</i> , intending to not <i>A</i> }
Cyclic preferences:	{preferring <i>A</i> to <i>B</i> , preferring <i>B</i> to <i>C</i> , preferring <i>C</i> to <i>A</i> }
Inter-level incoherence:	{believing that one's evidence decisively supports <i>p</i> , not believing <i>p</i> }

In each case the set of attitudinal mental states includes at least one state (or subset of states) that structurally opposes another and none that support others, and so the sets count as incoherent. Contrast that with the following:

Deductive coherence:	{believing that <i>p</i> , believing that if <i>p</i> then <i>q</i> , believing <i>q</i> }
Means-end coherence:	{intending to <i>A</i> , believing that <i>A</i> -ing requires <i>B</i> -ing, intending to <i>B</i> }
Acyclic preferences:	{preferring <i>A</i> to <i>B</i> , preferring <i>B</i> to <i>C</i> , preferring <i>A</i> to <i>C</i> }
Inter-level coherence:	{believing that one's evidence decisively supports <i>p</i> , believing <i>p</i> }

¹³ Cf. also the distinction between "weak" vs. "strong" coherence in Section I.

In each case the set of attitudes includes at least one attitude (or sub-set of attitudes) that structurally supports another and none that oppose others, and so the sets (correctly) counts as coherent.

What about sets of attitudes that are neither coherent nor incoherent? The pressure-based view readily accounts for them. Consider, for instance:

Belief-consistency:	{believing that <i>p</i> , believing that <i>q</i> }
Intention-consistency:	{intending to <i>A</i> , intending to <i>B</i> }
Belief-intention consistency:	{believing that <i>p</i> , intending to <i>A</i> }
Intention-hope consistency:	{intending to <i>A</i> , hoping to <i>B</i> }

At least considered on their own, in the abstract, such combinations of (merely) consistent attitudes are neutral—they neither clash nor positively cohere. And that’s just what the pressure-based view predicts: none of the attitudes involved oppose or support the other.¹⁴ Consistency may be necessary for coherence, but it’s not sufficient.

Notably, we can construct a Worsnip-style test for when an attitude (or set of attitudes) structurally supports—i.e. generates structural support or pressure for—another:

Support Test	A set of attitudes <i>S1</i> structurally <i>supports</i> an attitude <i>S2</i> iff it is (partially) constitutive of (at least some of) the attitudes in <i>S1</i> that any agent who holds <i>S1</i> has a disposition, when all of the attitudes in <i>S1</i> are occurrent (rather than merely dispositional), to have <i>S2</i> . ¹⁵
---------------------	--

To illustrate: it’s plausibly constitutive of the attitudes involved that if you occurrently believe that *p* and that if *p* then *q*, then you’ll be (strongly but defeasibly) disposed to believe that *q*, and if you occurrently intend to *A* while occurrently believing that *A*-ing requires *B*-ing then you’ll be disposed to intend to *B*. Similarly, it’s plausibly constitutive of the attitudes involved that if you occurrently believe that *X* is dangerous then you’ll (again, defeasibly) be disposed to fear *X*, and if you occurrently prefer *A* to *B* and *B* to *C* then you’ll be disposed to prefer *A* to *C*. This aligns with the pressure-based verdicts: believing that *p* and that if *p* then *q* structurally supports believing that *q*, intending to *A* and believing that *A*-ing requires *B*-ing supports intending to *B*, believing that *X* is dangerous structurally supports fearing *X*, and preferring *A* over *B* and *B* over *C* structurally supports preferring *A* over *C*.

¹⁴ This claim is intended as a generic, and hence compatible with there being exceptions, depending on the specific content of the relevant states.

¹⁵ I’m assuming *S2* is not included in *S1*. Also, I’m following Berker ms in appealing to conditions of “simultaneous occurrency” rather than Worsnip’s notion of full transparency—cf. footnote 15.

We can construct an analogous Worsnip-style test for when an attitude (or set of attitudes) structurally opposes—i.e. generates structural pressure *against*—another:

Opposition Test A set of attitudes S1 structurally *opposes* an attitude S2 if and only if it is (partially) constitutive of (at least some of) the attitudes in S1 that any agent who holds S1 has a disposition, when all of the attitudes in S1 are occurrent, to not have S2.

To illustrate: it's plausibly constitutive of the attitudes involved that if you occurrently believe that *p* and that if *p* then *q*, then you'll be disposed to not believe that not-*q*, and if you occurrently intend to A and occurrently believe that A-ing requires B-ing then you'll be disposed to not intend to not-A. Similarly, it's plausibly constitutive of the attitudes involved that if you occurrently believe that X is not dangerous then you'll be (defeasibly) disposed to not fear X, and if you occurrently prefer A to B and B to C then you'll be disposed to not prefer C to A. Again, this aligns with the pressure-based judgments: believing that *p* structurally opposes believing that not-*p*, intending to A together with believing that A-ing requires B-ing structurally opposes intending to not-A, believing that X is not dangerous opposes fearing X, and preferring A over B and B over C structurally opposes preferring C over A.

As with the earlier tests for (in)coherence, Support Test and Opposition Test can both be modified to allow for degrees by appealing to differences in the strength of the relevant dispositions.¹⁶ And (again) as with the earlier tests, Support Test and Opposition Test allow for a meaningful tripartite distinction between support, opposition, and the absence of both. For example, it's not constitutive of any attitude in any of the sets of consistent attitudes listed above (Belief-consistency, Intention-consistency, etc.) that anyone who has the relevant attitude (whether occurrently or not) has a disposition to either have or not have the other attitude in the set. Or, to make things slightly more concrete, consider the following sets of attitudes:

A1 = {believing that it's going to rain, preferring vanilla to chocolate}

A2 = {intending to watch the Lakers game}

It's not plausibly constitutive of the attitudes involved that anyone who has A1 (even when occurrent) has a disposition to either have or not have A2, nor that anyone who has

¹⁶ They can also be naturally extended to provide tests for when one attitude (or group of attitudes) *coheres with* one another and when an attitude (or group of attitudes) *clashes with* another—the two relational (in)coherence-like properties mentioned earlier. The extension is straightforward on the assumption that (a) S1 coheres with S2 just in case either S1 supports S2 or S2 supports S1 and (b) S1 clashes with S2 just in case either S1 opposes S2 or S2 opposes S1.

A2 has a disposition to either have or not have the attitudes in A1. So neither A1 nor A2 is predicted to support or oppose the other, which is the intuitively correct verdict.

Notably, Support Test and Opposition Test are broadly consistent with tests like Coherence Test and Incoherence Test, and if anything are more fundamental. For just as the coherence of a set of attitudes is (on the pressure-based view) a function of the more localized support relations obtaining within it, so the disposition to retain a coherent set of attitudes (if such there be) is plausibly a function of the more localized dispositions to form and/or maintain particular attitudes within the set. And just as the incoherence of a set of attitudes is a function of the more localized opposition relations obtaining within it, so the disposition to revise an incoherent set of attitudes is plausibly a function of the more localized dispositions to revise particular attitudes within the set.

4. An Objection and Reply

According to Worsnip, one of the main drawbacks of the pressure-based view is that it's committed to there being "ubiquitous" conflicts between substantive and structural rationality and that "in many cases in which one's initial attitudes are substantively irrational, structural rationality mandates one to compound these mistakes by adopting further, substantively irrational attitudes" (2021: 193).¹⁷ Worsnip thinks that a view that minimizes substantive-structural conflicts—such as his own wide-scope(-in-spirit), requirements-centric view—is preferable.

To illustrate the purported problem, let's return to the case of Tom, who believes that he's Superman and that Superman can fly but doesn't believe he himself can fly—indeed, he believes that he *can't* fly. Tom is clearly structurally irrational, as Worsnip and I both agree. We also agree that in order to be structurally rational Tom must revise one or more of his beliefs, with no belief in particular being singled out (structurally speaking) for revision. Whereas Worsnip appeals to the Incoherence Test and corresponding wide-scope requirement to explain Tom's case, however, I appeal to facts about structural pressure. In brief: any two of Tom's three beliefs structurally oppose the third—in general, $B(p)$ and $B(\text{if } p \text{ then } q)$ opposes $B(\text{not-}q)$, $B(\text{not-}q)$ and $B(\text{if } p \text{ then } q)$ opposes $B(p)$, and $B(p)$ and $B(\text{not-}q)$ opposes $B(\text{if } p \text{ then } q)$ —and the amount of opposition generated in case is roughly on a par. So although Tom must revise at least one of his beliefs in order to resolve the tension, the relevant facts about structural pressure don't favor any particular way of doing so. This is a case of what I call *attitudinal conflict* (Fogal 2020). In cases of attitudinal conflict, then, while the verdict of structural rationality doesn't align with that of substantive rationality (which, in this case, directs Tom to *not* believe that he's Superman and *not* believe that he himself can fly), it doesn't direct conflict with it either—the demands are co-satisfiable. The same is true on Worsnip's view.

¹⁷ Besides raising objections to the pressure-based view (in §6.6.2), Worsnip also replies to (some of) my objections to the requirements-based view (in §6.6.1).

Our views diverge, however, with respect to a slightly different version of the case. To illustrate, let's introduce another new character: Todd, the youngest brother of Tom, Tim, and Tam. Like Tom and Tim, Todd believes that he is Superman and that Superman can fly. But unlike Tom and Tim, Todd has *no attitude at all* towards the proposition that he himself can fly. Todd is a cognitively limited agent, after all, with finite resources, and so (let's suppose) he simply hasn't yet 'put two and two together'.

On Worsnip's view, Todd is structurally irrational—his set of beliefs is incoherent (since it passes the Incoherence Test), and so is structurally prohibited. On my view, however, he's not.¹⁸ For while there is significant undefeated structural pressure for Todd to believe that he can fly—after all, Todd's standing beliefs (namely, that he's Superman and that Superman can fly) generate significant structural pressure to believe that he himself can fly, and there's no comparable structural pressure to revise them¹⁹—I don't think Todd is automatically structurally irrational in virtue of not (yet) having considered the relevant question. We're not structurally required, after all, to have deductively closed beliefs (as Worsnip himself agrees²⁰), nor are we required to have every attitude there's undefeated support for.²¹ At most, then, we can say that, given his existing beliefs, Todd *conditionally* ought to believe that he can fly, where the 'ought' is conditional in two ways: it's *conditioned* on the existence of Todd's standing, structural-support-generating beliefs as well as *conditional* on Todd considering, or otherwise beginning the process of forming an attitude towards, the proposition that he himself can fly.²²

It's with these sorts of cases—i.e., cases *not* involving attitudinal conflict—that Worsnip's worry arises. On my view, given that there's significant structural pressure to believe that he (himself) can fly and no opposing structural pressures, Todd structurally (albeit conditionally) ought to believe that he himself can fly—that's the all-things-considered structurally favored response. Substantively speaking, however, Todd ought *not* to believe that he can fly—his evidence supports believing he *can't* fly. Is this "conflict" between substantive and structural rationality problematic? I confess to being unmoved. On the contrary, it seems intuitively correct—Todd's existing beliefs *do* seem to structurally favor (without requiring) a particular response, while his evidence singles out the opposite response as being substantively (and rather decisively) favored.

Granted, this might seem to boil down to little more than registering a difference in intuitions. Fortunately, more can be said. For it's worth noting that Worsnip doesn't deny

¹⁸ What I say here is compatible with but goes beyond what I say in Fogal 2020.

¹⁹ That's because absences of attitudes (in this case, the absence of the belief that he can't fly) don't generate any structural pressure—only attitudes themselves do.

²⁰ It may be a rational *ideal* but it's not a requirement. For Worsnip's argument against the deductive closure requirement, see §9.4 (especially pp. 298-301).

²¹ Cf. Nelson 2010, whose argument that we have no positive epistemic duties arguably generalizes.

²² I take verdicts involving so-called "weak" deontic necessities (e.g. 'ought (to)', 'should') to be more plausible than "strong" ones (e.g. 'have (to)', 'must') in cases like Todd's, since they better capture, in modal form, judgments concerning what the structurally favored—and hence "best"—response is.

that structural and substantive rationality can conflict—on the contrary, he argues positively that they can (see §5 of Chapter 3). He just doesn't think they conflict very often. But given dualism about rationality, it's not clear why we should be any more bothered by regular conflicts between the verdicts of substantive and structural rationality than we are by, say, regular conflicts between verdicts of morality and self-interest, or between "objective" and "evidence-relative" deontic verdicts (i.e. what one ought to do, given all the facts, vs. what one ought to do, given one's evidence).

It's also worth noting that the very feature of Worsnip's view—namely, acceptance of the wide-scope(-in-spirit), requirements-centric account of structural rationality—that enables it to avoid positing widespread substantive-structural conflicts is also what prevents it from recognizing the possibility of widespread substantive-structural *alignment*. That's because substantive rationality will typically single out individual attitudes as required (or otherwise favored), while structural rationality won't. On the pressure-based view, however, the verdicts will often align: in many cases in which one's initial attitudes are substantively rational, structural rationality will favor the response(s) that lead to further, substantively rational attitudes. This is of course the positive flip-side of Worsnip's objection. The question thus arises: if minimizing *conflicts* between substantive and structural rationality is desirable, why isn't increasing *concord* also desirable? Worsnip's view scores well on the first metric, and poorly on the second. It's not clear why the former should be seen as preferable—or as desirable at all, given dualism.²³

5. Conclusion

To sum up: *Fitting Things Together* is a truly terrific book, consisting of an extended defense of dualism about rationality together with a novel, detailed development of a wide-scope, requirements-centric account of structural rationality. There's just one major oversight: it doesn't say anything about the difference between coherence (in the fitting-things-together sense) and the mere lack of incoherence. The pressure-based alternative that I favor, by contrast, promises to capture *both* coherence *and* incoherence *and* a range of other (in)coherence-like properties, and is otherwise compatible with many of the main claims and arguments of *Fitting Things Together*. All it requires is the rejection of the wide scope, requirements-centric picture—a price we should be independently happy to pay.²⁴

²³ Worsnip raises a separate worry about the possibility of widespread structural-substantive conflicts, claiming that if the verdicts of structural rationality regularly conflict with those of substantive rationality then "it seems dubious that it has much, if any, normative significance", where 'normative significance' is cashed out in terms of normative reasons (2021: 177). My view is officially neutral on the normative significance of structural rationality, though as a dualist I'm fine with the lack of any essential connection between reasons and structural rationality. Both Worsnip and I agree, however, that not everything of significance *simpliciter* is of specifically *normative* significance, and in particular that structural rationality needn't be normatively significant in order to count as a genuine dimension of evaluation. See Worsnip 2021: 29-30 for elaboration.

²⁴ Thanks to Selim Berker, Max Hayward, Benjamin Kiesewetter, Olle Risberg, and Keshav Singh for helpful feedback on an earlier draft of this paper. A special debt of gratitude is owed to Alex

Bibliography

- Berker, S. ms. Worsnip's Unification of Structural Rationality. Talk delivered at the Orange Beach Epistemology Workshop, May 2022.
- Broome, J. 2013. *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.
- Brunero, J. 2020. *Instrumental Rationality: The Normativity of Means-Ends Coherence*. Oxford: Oxford University Press.
- Fogal, D. 2018. On the Scope, Jurisdiction, and Application of Rationality and the Law. *Problema* 12: 21–57.
- Fogal, D. 2020. Rational Requirements and the Primacy of Pressure. *Mind* 129(516): 1033–1070.
- Fogal, D. & Risberg, O. ms. Coherence and Incoherence.
- Fogal, D., & Worsnip, A. 2021. Which Reasons? Which Rationality? *Ergo* 8: 11.
- Kiesewetter, B. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.
- Lord, E. 2018. *The Importance of Being Rational*. Oxford: Oxford University Press.
- McGrath, M. 2021. Epistemic Norms on Waiting (and Suspending). *Philosophical Topics* 49(2): 173-202.
- Nelson, M. 2010. We Have No Positive Epistemic Duties. *Mind* 119(473): 83-102.
- Pryor, J. 2004. What's Wrong with Moore's Argument? *Philosophical Issues*, 14: 349–378.
- Pryor, J. 2018. The Merits of Incoherence. *Analytic Philosophy*, 59(1): 112–141.
- Way, J. 2018. Reasons and Rationality. In D. Star (ed.), *Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press.
- Wedgwood, R. 2017. *The Value of Rationality*. Oxford: Oxford University Press.
- Worsnip, A. 2021. *Fitting Things Together: Coherence and the Demands of Structural Rationality*. Oxford: Oxford University Press.
- Von Wright, G. H. 1963. *Norm and Action: A logical inquiry*. London: Routledge & Kegan Paul.

Worsnip, not only for his detailed comments but also—and more importantly—for the many hours of stimulating conversation we've had about rationality (among other things) over the past ten years.