BOOK REVIEWS

Fitting Things Together: Coherence and the Requirements of Structural Rationality,
Alex Worsnip, Oxford University Press, 2021, xvii + 335 pages
doi:10.1017/S0266267123000226

In this superb, densely argued book Alex Worsnip presents an account of the distinctiveness of structural rationality that seeks to explain what makes it, and the requirements it imposes, unified, distinctive and normative. At its heart is the view that structural rationality is a property of *sets* of mental or propositional attitudes that is absent when the elements of the set in question do not 'fit together' coherently. Familiar examples are a set of cyclical preferences or the set of beliefs {A, If then B, Not B}. This is in contrast with substantive rationality, a property of singleton attitudes that tracks their responsiveness to reasons, i.e. to objective features of the world.

The territory intersected by economics and philosophy abounds with claims about what rationality requires of agents, but much of the debate around them is 'naïve' in the sense that it operates with a vague concept of rationality and, in particular, without much attention to the distinction between structural and substantive rationality. Consider for instance the lively and ongoing debate in decision theory over the status of the Sure-thing principle or in formal epistemology over that of conditionalization as a rule of probabilistic belief revision. Both have centred on whether rationality requires conformity with these rules, but without much clarity as to what kind of rationality is at stake and what its normative significance is. Worsnip's book offers a way of answering these questions in a principled way. This alone should make it of interest to many readers of this journal.

Fitting Things Together is divided into three parts. In the first, Worsnip defends his 'dualist' position that structural rationality is distinctive from substantive rationality, by evaluating and rejecting the various claims of reducibility or eliminability – of structural rationality to substantive rationality, or vice versa – to be found in the literature. The second part develops his positive account of structural rationality, including of what unites instances of it, what requirements it imposes and what makes them normative. In the third and final section he draws on this account to shed light on a variety of other philosophical issues, including moral rationalism, the nature of rational choice theory and the normativity of logic. There is a lot going on in this last section, but I will have little to say about it. Instead the focus of my attention will be the positive account of structural rationality developed in the second part of the book and to a lesser degree the argument for dualism and the irreducibility of structural rationality that he makes in the first part.

Let's start with Part II, where Worsnip gives the core positive account of structural rationality. In Chapter 5, we get a characterization of incoherence as a property of sets of attitudes and a corresponding claim about what structural rationality requires of us; in Chapter 6, an account of what requirements structural rationality imposes on us and in Chapter 8, an explanation of the normativity of structural rationality. The two core claims underpinning this account are that:

Incoherence: A set of attitudes (or attitudinal states) is incoherent iff it is partially constitutive of these attitudes (or states) that an agent who holds them jointly will be disposed under conditions of full transparency to revise at least one of them.

Rational Requirements: The requirements of structural rationality are prohibitions on the adoption of incoherent sets of attitudes.

Much, of course, depends on the details: on what the conditions of 'full transparency' are, what counts as possessing an attitude and what it means for a disposition of the relevant kind to be constitutive of an attitudinal state. Notably, on Worsnip's account, full transparency implies knowing what one believes – something that he explicitly denies is our normal state. So I could believe that P, that if P then Q and that ¬Q, without being structurally irrational, just so long as I am not aware that I hold all of these beliefs. But even this leaves open space for different 'grades' of structural irrationality depending on whether the beliefs that must be fully transparent are just the active ones (i.e. concern propositions to which one is attending at the time) or whether they include those that are not currently active but that one could recall or that could come into attention or indeed those that one would form if one became aware of the possibilities they concern.

These details aside, on the face of it *Incoherence* provides for a compelling unified characterization of structural rationality. And tying coherence to conditions constitutive of a type of attitude has some interesting implications. Firstly, it offers the possibility of a characterization of the different propositional attitudes in terms of the revision dispositions that pick out the associated coherence properties. And secondly, it offers an explanation for why it is that concerns about the requirements of structural rationality (rather than substantive rationality) are and should be central to the enterprise of modelling and explaining behaviour and choice and not just to its rationalization. For if it is partially constitutive, for example, of being in a state of *believing* that P, that one is disposed, in conditions of full transparency, to revise any set of attitudes to which it belongs that also contains the belief that one does not believe that P, then one could not attribute to someone both the belief that P and belief that they do not believe that P, on pain of misidentifying the attitude of which P is the content. Note that the presumption of structural rationality does not imply that one could never ascribe incoherent beliefs to an agent, because all that is required of them is that they are disposed to revise any such incoherent set of attitudes in relevant circumstances. So ascription of incoherence can proceed

when accompanied by a hypothesis that the conditions of full transparency do not hold. In this respect Worsnip's theory improves on standard versions of interpretivism which render the identification of irrational attitudes somewhat mysterious.

Jointly *Incoherence* and *Rational Requirements* tell us that one ought not, on pain of structural irrationality, be in a mental state such that anyone in it would, under conditions of full transparency, be disposed to revise it. Note that the requirements in question are not 'narrow-scope'; that is, they don't oblige an agent, for instance, to adopt the belief that Q if they believe that P and that if P then Q. Rather they are 'wide-scope' in that they relate to sets of attitudes: they, for instance, prohibit the agent from adopting the triple of beliefs {P, if P then Q, Q}, a prohibition that doesn't rule out adoption of the belief that Q.

This has interesting consequences for many of the debates about the principles of rationality to be found in the economics and philosophy literature. For instance, it follows from it that Completeness is not a principle of structural rationality since it is false to say that everyone is disposed to revise incomplete sets of attitudes even under conditions of full transparency. The principle of Transitivity is an even more interesting test case. If it is a condition of structural rationality, then it must be read as prohibiting states containing a preference for P over Q, for Q over R and for R over P. But it *cannot* be read as the requirement on anyone who prefers P over Q and Q over R, to prefer P over R. This latter, narrow-scope requirement, can only be obtained in conjunction with two other requirements (1) to adopt some preference between P and R, and (2) to retain current preferences between P and Q and between Q and R. Neither are plausibly generated by *Incoherence*.

This is puzzling in some ways, because we are often inclined to invoke Transitivity to derive narrow-scope requirements of this kind. Much of Chapter 7 is devoted to developing a broadly contextualist account of the way in which coherence conditions on attitudes are deployed in deliberation, an account which offers an interesting response to this problem. The essence is that the context of deliberation fixes certain facts upon which the requirements of rationality are conditioned. For example, in a context in which it is given that I prefer P to Q and Q to P, then the only preference between P and R that I can consistently adopt is for P over R. That is not of course the same as saying that I should prefer P over R – for that it would be necessary to build (1) into the context as well. Another way of putting this is that the wide-scope principle of Transitivity yields in contexts containing (1), but not (2), the narrow-scope requirement that preferences be Suzumura-consistent, and in contexts containing both (1) and (2), the stronger narrow-scope requirement that preferences be transitive. Worsnip's account thus entails that Suzumura-consistency and Transitivity are not different coherence conditions but different narrow-scope requirements generated by the same underlying principle of structural rationality, depending on whether or not the context builds in the requirement of Completeness.

Worsnip's view also has implications for debates around diachronic requirements, such as that of conditionalization. Bayesian conditionalization requires of agents that, in the event of learning the truth of evidence proposition E, they adopt a degree of belief in any other proposition Q equal to their initial conditional degree of belief in Q given that E. There is a wide-scope

requirement nearby; namely the prohibition on adopting probabilistic degrees of belief Bel such that for all x y, $Bel(E) = 1$, $Bel(Q/E) = y$ and $Bel(Q) = x$. But this requirement can be met by revising $Bel(Q/E)$ or even $Bel(E)$, so it does not suffice to generate the narrow-scope requirements on posterior degrees of belief that Bayesianism is typically seen as imposing. Now I think it's reasonably plausible that the rigidity of one's conditional degrees of belief given the evidence can be taken to be a standard part of the context in which conditionalization applies (though this clearly needs filling out if one wants to derive a narrow-scope requirement to conditionalize). But, while it might well be a matter of context whether or not some proposition E is part of someone's evidence, it is surely not a contextual matter whether or not they take their evidence to be true (if anything this is a requirement of substantive rationality). The upshot is that Bayesian conditionalization cannot be regarded as simply a condition of structural rationality.

There is a lot more to be gleaned from Worsnip's positive account of structural rationality, but let me turn now to a different set of issues and to the contribution that the first part of his book makes to the question of what rationality consists in. As we have seen Worsnip defends a dualist view on which there are two distinct and autonomous conceptions of rationality. While structural rationality requires that one's attitudes fit together coherently, substantive rationality requires that one's attitudes respond to the reasons one has. In the early part of the book Worsnip offers both his own account of what substantive rationality consists in and rebuttals of the various arguments to be found in the literature that one or other conception of rationality is primary and that the other can either be reduced to it or eliminated altogether.

I found Worsnip's case against the reduction or elimination of structural rationality unassailable. To get a flavour, consider an argument frequently put forward for the eliminability of structural rationality. This thought is that the requirements of substantive rationality on their own suffice to ensure that someone's attitudes will cohere, because if each element of a set of attitudes adequately responds to the reasons they have, then these attitudes cannot (by the nature of the reasons they respond to) contradict each other. But, as Worsnip points out, whether this is so or not, it is still the case that someone could satisfy the requirements of structural rationality even if their attitudes are not substantively rational e.g. if they adopt a consistent set of false beliefs. So it doesn't suffice to be substantively irrational to be structurally so; hence the requirements of structural rationality are autonomous.

A second argument for elimination starts with the thought that requirements of coherence don't generate reasons for taking attitudes. For instance, the belief that P is not a reason for believing that P. The reason to have this belief (if there is one) lies outside of one, in the fact that P or in the existence of evidence that P. Now this observation is not something Worsnip needs to reject for, as we have seen, he does not claim that structural rationality entails narrow scope requirements to take attitudes. Indeed it's a feature of his account that such a narrow scope requirement to believe that P in virtue of believing that P could only arise in a context in which P (or the belief that P) is taken for granted. One could nonetheless still see these contextual narrow scope requirements as giving reasons for holding an attitude which are subjective or internal in that they

arise, in particular contexts, from the attitudes the agent already holds. But the essential point is that one does not generally satisfy the requirements of structural rationality simply by adopting the attitudes that one has reason to adopt. That one believes that P may or may not give one reason to believe that P, but this fact is quite distinct from the requirement to not believe *both* P and ¬P, a requirement that could be violated by someone who correctly responds to the reason that have to believe that P (or ¬P, as the case may be).

To support the other half of his dualism, Worsnip needs to give some account of substantive rationality qua responsiveness to reasons. The challenge here is there are different kinds of reasons that one might respond to, and because what attitudes these different reasons support might well be different, something must be said about which class of reasons one must respond to (and which one must not) if one is to count as substantively rational. Worsnip in particular distinguishes between responsiveness to fact-relative, evidence-relative and belief-relative reasons. Substantive rationality cannot, he argues, require responsiveness to the facts, because this would make it implausibly demanding. On the other hand, mere responsive to one's beliefs does not suffice. Responsiveness to one's evidence is therefore his proposed Goldilocks requirement: not too demanding, not too permissive, just right.

I am not entirely convinced by this claim. If one's evidence is simply what one knows or believes with certainty then responsiveness to one's evidence reduces to responsiveness to one's true (full) beliefs. This seems too weak, because this kind of responsiveness to the evidence is consistent with adopting false beliefs not supported by the evidence alone but which can be inferred from the evidence together with other false beliefs one holds (indeed, sometimes must be inferred on pain of structural irrationality). On the other hand, if one can have evidence that one does not believe true or is not aware of, then requiring responsiveness to evidence seems implausibly demanding in much the same way as requiring responsiveness to the facts. For example, suppose that someone's testimony to the effect that the trains are delayed gives me reason not to make a train trip, but that I don't believe the testimony. Then it doesn't seem right to say that it is irrational for me to attempt the trip even if I have reason not to. Holmes may have been correct in taking the dog's not barking as evidence that the murderer was no stranger, but Watson is not irrational for failing to recognize this.

Perhaps one could respond to this challenge by grasping the first horn of the dilemma but insist that one can know the evidence while still failing to recognize what is supported by the evidence, so that the violation of substantive rationality lies in not having the right beliefs of the E-entails-that-F kind. But why is it irrational not to know what is supported by one's evidence? Suppose it's true that a liquid smelling of almonds is evidence for it containing arsenic and consider the following three cases:

1. The liquid contains arsenic, but there is no evidence that it does and, indeed, you don't believe it does.
2. The liquid contains arsenic. Furthermore, it smells of almonds and the smell of almonds evidentially supports it containing arsenic. But although you believe it smells of almonds you don't believe that its smelling of almonds supports it containing arsenic. And, indeed, you don't believe that it does contain arsenic.

3. The liquid contains arsenic. You know that it smells of almonds and that the smell of almonds evidence evidentially supports it containing arsenic. You don't believe you have any other evidence to the contrary, yet you don't believe it contains arsenic.

In all three cases you have a fact-relative reason to believe that the liquid contains arsenic but don't in fact believe it to be the case. What kind of irrationality, if any, is involved in each case? It seems clear enough that case 1 does not involve irrationality of any kind and that case 3 is one of structural irrationality since your beliefs do not fit together. But though you fail in case 2 to respond both to your fact-relative reasons for belief and, on this proposal, your evidence-relative reasons, it is not clear to me that this makes you irrational. But perhaps this is getting close to a quibble about words.

Richard Bradley ⊙
London School of Economics and Political Science, London, UK
Email: R.Bradley@lse.ac.uk

Richard Bradley is Professor of Philosophy at the London School of Economics and Political Science and a Fellow of the British Academy. He works mainly on issues having to do with uncertainty and individual and social decision making, but has broad interests in the philosophy of economics and social science, in formal epistemology and in the semantics of conditionals.

The Pursuit of Happiness: Philosophical and Psychological Foundations of Utility, Louis Narens and Brian Skyrms. Oxford University Press, 2020, 208 pages.
doi:10.1017/S0266267123000202

Luis Narens and Brian Skyrms's *The Pursuit of Happiness* is an attempt to take seriously utilitarianism's problem with how to measure happiness. The problem is to find a way to make sense of measurements of happiness so that the utilitarian aggregates of happiness will be meaningful. The first part of the book provides a very useful historical overview of the measurement of happiness in utilitarian theory. As far as we know, it is the first overview of this kind. This part covers Bentham, Mill, Jevons, Edgeworth and the 19th-century psychophysics. It then goes on to present von Neumann and Morgenstern's method of generating cardinal utilities from ordinal preferences over lotteries and Harsanyi's aggregation theorem.[1] The second part provides an overview of modern measurement theory. This overview includes a discussion of modern

---

[1]The presentation of von Neumann and Morgenstern's expected-utility theorem has some unfortunate errors, however. On page 66, Ordering is defined as just completeness. But Ordering is completeness and transitivity, which is what is required for von Neumann and Morgenstern's theorem. (Likewise, the explanation of Ordering on page 67 makes the same mistake.) On page 67, in the definition of Independence, 'for all a' should be 'for all $a > 0$'. Otherwise, Independence would demand that p is preferred to p' only if something is preferred to itself.