

Book Review

Rationality and Belief, by Ralph Wedgwood. Oxford: Oxford University Press, 2023. Pp. x + 298.

This book is the second in a trilogy that Ralph Wedgwood has been writing on rationality. The first was the 2017 book, *The Value of Rationality*, about rationality generally; this book is specifically about epistemic or theoretical rationality; and the third will be about practical rationality. As in much of his work, Wedgwood works here on the big picture—there is a grand unified theory of rationality worked out across these three books, with this book giving the details of theoretical rationality, and this overall theory of rationality is itself part of a broader picture of normativity generally that Wedgwood gave in his 2007 book, *The Nature of Normativity*. Many philosophers working in the analytic tradition are not comfortable with this sort of grand-scale work. But fortunately, this book is organized in such a way that every aspect of the theory has its own chapter, and each section of each chapter gives a single argument. This makes the book eminently usable for epistemologists and metaethicists who have fundamental disagreements with many of the big-picture aspects of the theory but still find some of the individual arguments valuable to engage with. Many parts of the view have been developed in standalone papers over the past twenty or so years, but here they are gathered in one place, and reorganized in a way that makes it clear how they all fit together.

I will spend much of this review summarizing the overall view, saying where in the book each part of the view is argued for. Some of these arguments depend on each other, while many others can be accepted or rejected somewhat independently of the overall picture. But I will conclude with some more general thoughts about analytic epistemology suggested by my reading of this book.

Wedgwood uses the phrase ‘belief-system’ for all the doxastic and epistemic attitudes that a person has. The overall picture that he draws of a belief-system makes major distinctions between theoretical and practical attitudes, and he differentiates between the propositional and doxastic rationality of each of these components.

The central aspect of a belief-system, for Wedgwood, is a set of ‘theoretical credences’, including both unconditional and conditional degrees of belief (ch. 5.6). This implicitly makes use of the conceptual resources of the agent to define a (finite) set of propositions that the agent has attitudes to and a (finite)

set of epistemic possibilities these are over (ch. 5.2). There is a natural way to measure these attitudes on a numerical scale from 0 to 1 (ch. 5.5), and when they are perfectly propositionally rational they will be precise (ch. 8.6), they will satisfy the axioms of probability theory (ch. 8.3), and the conditional degrees of belief will perfectly satisfy the ratio analysis with respect to the unconditional (ch. 8.5). However, Wedgwood notes that we generally don't have the resources to ensure that our theoretical degrees of belief are always simultaneously perfectly precise and rational, so they are sometimes imprecise (ch. 5.7), and may not even satisfy the probability axioms. Moreover, even when perfectly precise, they are often undefined for some propositions constructed out of ones that we do have theoretical credences in (ch. 5.4).

Another aspect of a belief-system, for Wedgwood, which he spends less time discussing in this book, is the set of 'practical credences' that arise in various contexts, when the agent is faced with a decision problem. These practical credences (which are said to be the subject of the third book in this series that he is planning to write) must have some features that ordinary agents generally won't be in a position to rationally have for their theoretical credences (ch. 6.5). In particular, he argues that there must be finitely many epistemic possibilities that get all the probability, even though agents generally know they are not in a position to rule out all alternatives (ch. 6.6). Thus the agent will be disposed, in normal circumstances, to treat certain sets of possibilities as though they were certain, despite not being theoretically certain of them. These dispositions are what Wedgwood identifies as 'outright beliefs,' of the sort that epistemologists have been discussing for centuries (ch. 6.5).

The rational analysis of these practical credences, including the beliefs that give rise to them, is in large part left for the next book. But Wedgwood gives a detailed picture of rationality for theoretical credences. The fundamental evaluative feature of a credence is correctness (ch. 7). The highest possible credence (which we represent with the number 1) is the maximally correct one to have in a true proposition, and the lowest possible credence (which we represent with the number 0) is the maximally correct one to have in a false proposition. All other credences are more or less incorrect depending on how far they are from this maximally correct one, with distance measured by the square of the difference between the truth-value and the numerical probability representing the credence (ch. 8).

While correctness is the fundamental evaluative feature of a credence, the subject of this book is rationality. Correctness inescapably involves an external component in the actual truth of a proposition, while rationality is, for Wedgwood, an internal affair (ch. 1). Like some epistemologists (but not others), Wedgwood treats justification and rationality as synonyms (ch. 0), and he treats this as another normative feature, which comes in propositional and doxastic varieties (ch. 2). Correctness is the fundamental actual evaluative concept (ch. 8.1), but a proper evaluation of it requires access to the facts, which are not

generally available. Propositional rationality is the best possible internal evaluation of correctness, given everything that the agent has to work with—for this concept he has both a liberal reading of ‘possible’ and a distinctive view of what is ‘given’ to the agent. (This concept of ‘givenness’ is more about what determines the bounds of practical possibility than it is like the standard epistemologist’s concept of ‘evidence’, though it also ends up playing a similar role to evidence.) Doxastic rationality, by contrast, is a matter of how effective one’s dispositions are at bringing about an overall belief-system that is usually highly propositionally rational. Most of the book is about the propositional rationality of theoretical credences.

Propositional rationality is described by means of an extended metaphor of an ‘evaluating angel’. This angel does the best job possible of evaluating the agent’s theoretical credences, making use of nothing other than what is forced on to the agent. Because it is the best possible, it treats all *a priori* facts as certain (ch. 11.2.1), and because it is forced to deal with the agent as they are, the angel also treats facts about the agent’s own belief-system as certain (ch. 11.2.2). Because of the inferential role played by the concept of chance, the angel satisfies a version of David Lewis’s ‘Principal Principle’ (ch. 12.5). But most importantly, because doxastic states are inherently ‘sticky’, and are inherently aimed at the truth, the angel tries to evaluate the agent in line with the ‘shadow’ of the agent’s own prior credences, at least as much as is compatible with these more central idealizations (ch. 13).

These ‘shadow’ priors are among the more distinctive features of Wedgwood’s theory. They are what prevent the theory from being a very permissive one (of the sort I tend to favour, but which he rejects in chapter 12.3) without making use of Carnapian formalism or Williamsonian externalism to tie things down (ch. 12.4). The idea is that although an agent’s past credences are not actually present, they still (in normal cases) exert causal influence on the present, and a well-functioning agent will have dispositions to form credences that are shaped by these past credences. But Davidson’s Swampman doesn’t actually have these shadow priors (ch. 13.1), and features of the agent’s own past credences can be prevented from playing this role by structural features of the case, as in Frank Arntzenius’s ‘Shangri-La’ case (ch. 13.2).

Although there is much of this overall picture I disagree with, one point that I completely agree with is the argument that the concept of ‘evidence’, as traditionally applied by epistemologists (whether internalist evidentialists like Feldman and Conee, or externalists like Timothy Williamson, or orthodox subjective Bayesians), can’t pick out anything that plays the roles they want (ch. 4). I can’t do justice to the full argument here, but there are two main parts. One part involves the role that evidence is supposed to play—on all these views, evidence has propositional content, it is what drives all other credal states (either by uniquely determining the correct states to have or by shaping the update from *a priori*), and its content ends up having the highest credal status. But he argues that neither the propositional content nor the maximality of the resulting credal

state can be justified in general for the things that are supposed to play the role of evidence in particular cases.

The second part of the argument involves identifying the origin of the philosophical concept of evidence. He argues that the word 'evidence', in ordinary English, identifies whatever means have been deemed appropriate in a particular dialectical context (like a courtroom or a scientific journal) for the social justification of belief. Philosophers have mistakenly assimilated the private structure of our entire doxastic state with the public context of using a shared basis to come to agreement on a few potentially controversial beliefs. (He also argues that in other languages, like French and German, the closest cognate of the English word 'evidence' doesn't pick out facts of a special kind, but rather refers to the 'evident-ness' of a given fact.)

Nevertheless, there is a part of his account that does some of the work that evidence is supposed to on other accounts. The evaluating angel works with a 'shadow prior', which is an idealized version of the agent's own past credal state, and conditionalizes it on the 'given' features of the situation that are outside the agent's control but present in a way that can be used to shape the agent's current belief-system (ch. 9.2). This includes features like the existence of the agent, their immediate past, the output of mental modules that are sub-personal, like perceptual and memorial systems, and so on, but also includes the fact that the agent actually has the belief-system that they have (ch. 9.4). With this conditionalized function, the angel then estimates the expected accuracy of the agent's belief-system (since this is the one fundamental epistemic value), and uses this to measure the agent's degree of propositional rationality. The most rational belief-system will be the one that matches the angel herself, because the measure of correctness is a 'strictly proper scoring rule' (ch. 8.3), and other possible credences will have lower rationality, measured by the squared difference of the credences (ch. 9.6). If the agent's credences are imprecise, they are evaluated on average, and this ends up meaning that imprecise credences are always less propositionally rational than the precise credence that is their midpoint (ch. 10.2).

The angel is objectively doing the best she can do by taking into account the agent's past credences. This is said to follow from the same sort of arguments that lead to doxastic conservatism, namely that it is the nature of belief to be 'sticky', and there is no better available guide to the empirical truth than one's recent belief-system (ch. 13.5). Interestingly, part of this recent belief system is a large number of conditional credences, indicating one's inferential dispositions (ch. 5.6). Among these conditional credences are various credences about how the world is, conditional on the deliverances of one's sense and memory modules (ch. 14.2). As a result, Wedgwood is able to argue that one is usually propositionally justified in being highly confident in various propositions on the basis of sensory experience (if one is a normal human who usually has such high confidence), but that this isn't usually warrant for certainty or for conditionalizing on any proposition other than those about oneself. Furthermore, in

circumstances where one has come to doubt one's senses or memories, one is now propositionally justified in having lower credences in the propositions we usually describe as the contents of these experiences. This is an interesting and novel approach that seems to me to do a lot of the work that evidence was meant to do, but doesn't require postulating innate contents to any of these things.

Because there is no necessary connection between past credences and future credences, there is no unified 'ur-prior' that the shadow priors always come from (ch. 13.6). And both because the agent's own past credences are sometimes imprecise, and because the angel needs to conditionalize on propositions that the agent doesn't necessarily have credences about (such as facts about the agent's own mind) the shadow priors will often be imprecise (ch. 10.1). This causes an interpretative problem, because imprecise credences for the agent are always less than perfectly rational (ch. 10.2), while the angel must sometimes use an imprecise set. Awkwardly, this means that for a human the set of probability functions is treated as a representational convenience, while for the angel it is treated as an actual fact. He suggests some ways of using such a set to evaluate an agent, but they all lead to a kind of moderate permissivism about propositional rationality (that is, any credence function that some angelic credence estimates as maximally rational is ideally propositionally rational), rather than a requirement of imprecision.

Much less is said about doxastic rationality. In chapter 2, Wedgwood argues that correctness is primary, propositional rationality (that is, distance from the 'evaluating angel') is secondary, and doxastic rationality is tertiary. Doxastic rationality consists in having and manifesting cognitive virtues that tend to reliably produce credal states that are highly propositionally rational. In some cases, such as when an amateur mathematician is wondering about the digits of π beyond what they have memorized, the propositionally rational credences to have are all 1's and 0's, but the most reliable dispositions available to the person for the digits they haven't memorized will usually lead to credences of 0.1 in each possible digit. Thus these credences that are not propositionally rational may still be doxastically rational for the agent. There are some intriguing suggestions in chapter 11.3 that this distinction between propositional and doxastic justification may also explain what to say about cases of apparent higher-order defeat, such as cases of peer disagreement or hypoxia. He suggests that something like the 'right reasons' view is correct for propositional rationality, while allowing that something more conciliatory might be doxastically rational, but he doesn't give as many details as I would like. He makes a similar move to respond to cases of objectionable trade-offs in response to Selim Berker (ch. 7.6) and to cases of self-undermining belief (ch. 11.6).

In the very last section of the book, chapter 14.4, he finally discusses the concept of knowledge. Unlike many other philosophers who are inspired by the Bayesian tradition, he does treat knowledge as an important concept. But in his analysis, it comes last, and depends on all his other concepts, rather than coming

first and being a key source of insight into the others. Unlike most other contemporary epistemologists, Wedgwood even gives an attempted analysis of knowledge: knowledge is outright belief that is correct because it is rational. Because knowledge is a species of outright belief, it is an inherently practically-oriented concept (recall that outright belief, for him, is just a disposition to form practical credences of certainty in the proposition, in normal circumstances, when one needs certainties in order to rationally deliberate about how to act). The phenomena of pragmatic encroachment on knowledge are real, even though the propositional and doxastic rationality of theoretical credences have no pragmatic influences. Presumably, this practical connection means that there will be more on knowledge in the third book, perhaps including responses to the inevitable counterexamples that await any analysis of knowledge.

This is certainly an impressive achievement, to be able to produce a grand unified theory of this sort, and to have arguments justifying each part of the edifice. Although in my presentation, it seems that related issues are scattered throughout the book, his organization scheme is helpful for explaining the whole system. Chapter 0 gives his statement that rationality and justification are synonyms and that rationality is internal, and then outlines the book. Part I, consisting of chapters 1–4, motivates the general structure of correctness, propositional and doxastic rationality, and the challenge to the concept of evidence. Part II, consisting of chapters 5–8, gives the picture of credences, both precise and imprecise, theoretical and practical, and full belief, as well as the motivation for the fundamental use of a quadratic scoring rule to measure correctness. Part III, consisting of chapters 9–13, gives the whole theory of the ‘evaluating angel’ and ‘shadow priors’. Chapter 9 gives the basic idea, chapter 10 discusses technical issues about imprecision, and then chapters 11–13 argue for all the relevant features of these theoretical ideas. Finally, the Epilogue, consisting of chapter 14, gives the implication for various epistemological debates about defeat, perception, scepticism and knowledge.

But I have various worries about many of the starting points of the picture, many of which are worries that I have about a wide swath of work in analytic epistemology.

Once one has drawn the distinction between propositional and doxastic rationality or justification, why should one care about the propositional at all? On a certain kind of evidentialist point of view, evidence justifies beliefs, and this justification relation might be worked out through some sort of idealized inductive logic, or something of the sort. On that picture, if evidence and beliefs are propositions, then this justification relation is a relation among propositions that we might call ‘propositional justification’. But someone who has the relevant evidence and belief might not hold the belief on the basis of the evidence, and thus we need a separate theory of the ‘basing relation’ to understand when the belief itself is ‘doxastically justified’. But once one has given up the deontic conception of justification as an end in itself, and accepts that beliefs have an aim of truth (or practical success, or knowledge, or whatever else it might be),

why think that there is any interesting theory of propositional justification? The interesting theory would seem to be the one about doxastic justification: what methods can one use to ensure that one reliably ends up with correct beliefs (or successful, or knowledgeable, or whatever)? Aiming at propositional justification, which itself aims at correctness, seems like introducing an unnecessary intermediary. The beliefs that are output by the best methods (the propositionally justified ones) are only of situational interest — the methods themselves are what an agent should care about.

But more fundamentally, why should we think that one's overall system of beliefs must be decomposable into atomized attitudes like belief, suspension, and precise or imprecise credence, that take propositions as objects? Wedgwood states in chapter 5.1 that this is his starting point, but this is an unstated assumption in most contemporary analytic epistemology. Why not think that a doxastic state might be manifested as something much more inchoate? I think that this paradigm has been a side-effect of the symbolic model of computation, pushed in different ways by Turing and Chomsky, and the various forms of 'database' or 'language of thought' views of the mind that follow. But growing appreciation of the features of neural nets, and the fact that they seem no worse as a model of how the human mind works than these older models, may lead to a re-evaluation of this central paradigm of contemporary analytic epistemology. Such a more holistic model may often have features that are usefully reported in terms of individual attitudes to individual propositions, but there is no reason to think it is best decomposed into these attitudes, or evaluated by evaluating each individual attitude. It is only a database model that seems to me to make Wedgwood's finiteness assumption plausible — both Turing and Chomsky show that a finite symbolic system might have implicit dispositions towards an infinity of contents.

And this brings me to a final concern. Wedgwood forthrightly states at the end of chapter 0 that this book is a project in individual epistemology. An unstated assumption, though, seems to be that the individual is a human being much like you or me. This book is presented as an analysis of the concepts that we already use for thinking about belief and rationality, rather than a more open-ended theorizing about what sorts of concepts would be useful for radically different creatures. Just as there is no consideration of how groups could effectively process information in order to have reliable interactions with their environment, there is no consideration of how one might design an artificial intelligence, or how an animal or an alien radically different from ourselves might relate epistemically to its environment.

Some epistemologists may reject the goal of an internalist conception of rationality, or the unusual way that the doxastic/propositional distinction is drawn, or the way the 'given' shapes the 'evaluating angel'. But this is how I feel when I read many works of analytic epistemology that give a special role to knowledge, or that argue about whether belief is a type of credence or credence is a type of belief, or that argue that evidence gives the norms for belief. Many of

the framing assumptions of our different theoretical paradigms make our own work appear to be the needless exploration of baroque intricacies of an artificial system to other philosophers working in different paradigms. But if we do our work carefully, we can hopefully identify arguments and ways of thinking that will still be of use to those other philosophers, as this book is.

University of California, Irvine, USA
k.easwaran@uci.edu
<https://doi.org/10.1093/mind/fzae076>

KENNY EASWARAN