

Review of *The Value of Rationality* by Ralph Wedgwood

Benjamin Kiesewetter

Published in: *Ethics* 129 (3), 501–8 (2019), <https://doi.org/10.1086/701491>.

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

Sometimes, our attitudes, such as our beliefs and intentions, are incoherent in a way that seems to make it appropriate to call us irrational, which in turn suggests that having such attitudes violates certain norms of rationality. An intensely debated question in contemporary analytic philosophy – often discussed under the heading of the *normativity of rationality* – concerns the relevance of such norms and their violation. In his seminal article “Why be rational?” (*Mind* 114 [2005]: 509–563), Niko Kolodny raised the challenge to put forward a positive reason for complying with the norms of rationality. Famously, he argues that no plausible reason can be given, and that the assumption that rationality has normative significance is a myth. Not everyone has been convinced by his arguments, but many (including, most notably perhaps, John Broome, *Rationality Through Reasoning* [Chichester: Wiley Blackwell, 2013]) remain doubtful about whether the normativity of rationality can be vindicated.

It is very natural to read Ralph Wedgwood’s rich and sophisticated book *The Value of Rationality* as a contribution to the debate between Kolodny, Broome, and others, which aims to answer Kolodny’s challenge and to reject his skeptical conclusion. After all, according to what Wedgwood calls the “central claim” of his book, “the term ‘rational’ ... expresses a normative concept” (196). The first chapter presents four objections to this thesis. Three of these objections are dealt with in chapters 2 and 3, while answering the fourth – which is explicitly concerned with Kolodny’s challenge (38–39) – is the task Wedgwood takes up in “the whole rest of this book” (86). Caution should be exercised, however, since Wedgwood’s understanding of both rationality and normativity differs in important respects from that of other participants in the debate.

Firstly, Kolodny and Broome are concerned with a pretheoretical notion of rationality, which is anchored in ordinary judgments about what is rational or

irrational (cf. Kolodny 2005, 515). In contrast, Wedgwood is mainly concerned with a theoretical notion: he qualifies his central thesis by saying that “*when the term ‘rational’ is used in such branches of intellectual inquiry as formal epistemology and the theory of rational choice, it expresses a normative concept*” (196). As is well-known, in such theories the term ‘rational’ is often used in highly idealized ways that do not necessarily reflect ordinary judgments of rationality (for example when the theories assume that rational agents must be logically omniscient), and hence Wedgwood’s notion of rationality differs from Kolodny’s and Broome’s. This is why Wedgwood spends part of his book discussing objections to the normativity of rationality that did not worry Kolodny and Broome, such as that rational requirements based on an idealized notion of rationality cannot be normative because they do not seem to entail the agent’s ability to comply (see chapter 3). Wedgwood appears to assume, however, that his notion of ‘rationality’ is broader than the ordinary one that Kolodny and Broome focus on and thus incorporates the rational requirements they are concerned with. In this way, his project is of more general interest (in particular to formal epistemologists and decision theorists), and his thesis that rationality is normative is even more ambitious than the one that Kolodny rejected.

Moreover, while Broome and Kolodny aim at capturing ordinary judgments of rationality and irrationality in terms of purely structural norms of rationality, which govern the relations between what Scanlon has dubbed “judgment-sensitive attitudes” (attitudes such as beliefs and intentions), Wedgwood’s conception of rationality seems to include, in addition, all substantive internal norms of justification. For example, Wedgwood’s notion of rationality includes doxastic norms that are conditional on sensory experiences (cf. 12), such as those embraced by internalist foundationalists about justification – norms that others would consider to be requirements of substantive rationality (which may be understood in terms of responsiveness to available reasons) that need to be distinguished from requirements of structural rationality (or rationality as coherence).

Secondly, Broome, Kolodny and others (including, to put my cards on the table, myself in *The Normativity of Rationality* [Oxford: Oxford University Press, 2017]) adopt a reason-oriented picture of normative significance, according to which the view that rationality is normative can – and, indeed, should – be spelled out as the thesis that a

rational requirement to ϕ implies a (*pro tanto* or decisive) reason to ϕ . Wedgwood dismisses this approach. He even claims that “to achieve reliable insights into the normativity of the notion of rationality, ... talk of ‘reasons’ is best avoided” (87). Wedgwood presents his opposition to reasons as an objection to the so-called ‘Reasons First’ approach, according to which the notion of a reason is the most fundamental normative notion that can be used to explain other significant normative notions (see chapter 4). However, since understanding the normative significance of rationality in terms of reasons does not commit one to the ‘Reasons First’ approach, the rejection of this approach itself does not put such an understanding into question. As it turns out in the course of the relevant chapter, Wedgwood goes much further than denying that reasons come first – he denies that there is any important unified notion of a normative reason at all: “there are innumerable different concepts of ‘normative reasons’, none of them any more central than any other” (94).

What is the alternative picture of normativity that Wedgwood proposes? Generally speaking, Wedgwood’s approach might be described as putting *values first*: the “most fundamental” normative concepts are “the concepts of all the various kinds of values” (18). Wedgwood then characterizes the deontic in terms of the evaluative. A somewhat rough characterization of his contextualist account of ‘ought’ is that ‘It ought to be the case that A ϕ -s’ is true just in case A’s ϕ -ing maximizes expected value, relative to some kind of value and relative to some distribution of probabilities over possible worlds (chapter 5). Normative reasons, in turn, are understood in deontic terms (as explanations of ‘ought’ facts) as well as in evaluative terms (as ideal motivations or premises of good reasoning), with “normative-explanation reasons” and “ideal-motivation reasons” constituting different notions of normative reasons that are not coextensional (chapter 4). Hence, on Wedgwood’s picture, there is an “enormous variety of kinds of value” (15); each of these values generates, together with all the many possible probability distributions, a “great proliferation of ‘oughts’” (14); and this in turn gives rise (or is part of what gives rise) to an even greater proliferation of kinds of normative reasons.

As far as the normativity of rationality is concerned, instead of understanding it as the claim that rational requirements entail reasons, Wedgwood proposes to understand it as involving the following two principles: “(a) rational mental states ...

and rational mental events ... are in a way *good* ...; (b) if a thinker is rationally required to ϕ , then there is a sense in which the thinker ought to ϕ " (40). Since Wedgwood assumes in chapter 4 that a normative reason, in one sense, just is "a fact that contributes towards explaining a normative fact" (106), it follows – at least on the assumption that the 'ought' in principle (b) isn't entirely unexplained – that rational requirements entail normative reasons:

1. If you're rationally required to ϕ , then you ought to ϕ .
2. If you ought to ϕ , then there is some fact that contributes towards explaining that you ought to ϕ (assumption).
3. A fact that contributes towards explaining that you ought to ϕ is a reason for you to ϕ .
4. Therefore, if you're rationally required to ϕ , then there is a reason for you to ϕ .

Wedgwood does not mention that his view has this implication, presumably because he takes it to be a relatively uninteresting consequence of the more fundamental claims he makes about the value of rationality and the sense in which we *ought* to be rational. Nevertheless, given the well-known objections that Kolodny and others have raised against the particular thesis that rational requirements entail reasons, it is surprising that this implication is not defended or even discussed in this book. For example, the book contains no discussion of the worry that reason-entailing rational requirements would license unacceptable bootstrapping, nor of the difficulty of identifying a reason-giving fact that would count in favor of satisfying rational requirements. Wedgwood appears to think that he does not need to worry about these problems because he does not understand the normativity of rationality in terms of reasons. But since his view implies that rational requirements entail reasons, the problems seem to apply to it.

Let's have a closer look at how Wedgwood understands his two principles. Given the innumerable senses of 'ought' that his account generates, which one is the sense in which we ought to do what's rationally required? Wedgwood tells us that "the kind of 'ought' implied by rational requirements is a subjective or information-relative 'ought'" (61). This means that the relevant probability distribution, which is an essential

ingredient of every 'ought'-proposition according to Wedgwood's theory, is "in some way determined by the totality of the agent's mental states" (59).

But what is the value that determines the 'ought' of rationality? *Prima facie*, one might think that it should be overall value. However, this view runs into trouble because it generates what are sometimes called "reasons of the wrong kind" for attitudes, while Wedgwood accepts (along with many others) that rationality is insensitive to such reasons (52). For example, if having an epistemically unjustified belief promises certain benefits, this belief might be required by a subjective 'ought' that is relativized to overall value, but this does not seem to show that such a belief would be rational. Therefore, the 'ought' entailed by rational requirements cannot be the 'ought' that tells us to maximize expected overall value.

So what is the relevant value that determines the sense in which we ought to be rational instead? Wedgwood promises to answer this question in chapter 6, in which he puts forward the eponymous claim of his book, namely that "rationality is itself a kind of value, a way in which mental events or collections of mental states can be good" (136). According to Wedgwood, rationality is a value because rationality is a virtue – "broadly akin to the cardinal virtues of justice, wisdom, courage, and temperance" (200). Like these other virtues, rationality "involves three related kinds of goodness: i.) the goodness of a disposition; ii.) the goodness of the performances that manifest this disposition; iii.) the '*abstract*' goodness of the performances that this disposition normally produces" (141). These kinds of goodness are not relative to purposes or standards, but "non-relative or absolute" values (201). Against this backdrop, Wedgwood proposes to understand rational requirements, in analogy to other requirements of virtue, as claims about what is necessary to be as rational as possible (149-151). On this approach, a rational requirement is "a kind of 'must', not a kind of 'ought'" (150). However, given the axiological assumption that rationality is a value, it follows from Wedgwood's account of 'ought' that rational requirements are "necessarily coextensive with a kind of 'ought'" (151).

As far as I can see, however, this 'ought' is not the subjective or information-relative 'ought' that Wedgwood announced earlier in his book as constituting the 'ought' of rationality (see, e.g., 18 and 61). For an 'ought' that is implied by a requirement of virtue is one that tells us to maximize an *actual* value relative to the

omniscient probability function (cf. 150), not one that tells us to maximize some expected value relative to a limited epistemic standpoint. It thus turns out that Wedgwood's 'ought' of rationality is, according to his own account, a fully objective one. This is not by itself an objection to his view, but it is a point where the reader might find it difficult to grasp how Wedgwood's ideas about the 'ought' of rationality fit together and more guidance would have been helpful. Indeed, the subjective 'ought' that we have been promised in the beginning as the 'ought' of rationality is never elucidated in the book and never mentioned again after chapter 5 – a chapter that provides a detailed formal apparatus for understanding more or less subjective and objective 'oughts' in general, but no conception of the 'ought' of rationality in particular.

This is particularly surprising because later chapters do in fact contain the material to identify a subjective 'ought' that is entailed by rational requirements. In chapters 8 and 9, Wedgwood develops the idea that "rationality is not a completely *free-standing value*", but depends on its connection to "the value of correctness" (230). Correctness is the property that an attitude has when it satisfies a certain kind of external norm that applies constitutively to all attitudes of that type (or when it achieves its "external aim" as it is sometimes metaphorically described) (5). For example, "correctness for beliefs is a matter of the truth of the proposition that is believed; and correctness for choices is a matter of the feasibility and value of the course of action that is chosen" (242). The relation between rationality and correctness is supposed to answer what Wedgwood takes to be the most significant challenge for the view that rationality is normative, namely to explain what is good about the kind of merely internal coherence that rationality, according to him and many others, amounts to. In a nutshell, his answer is that achieving this coherence is the best way to pursue the external value of correctness, and in fact the only way to pursue this value in a guided way (235).

Now, in the course of chapter 9, Wedgwood maintains that "having correct attitudes ... is a genuine value" (231), and that "to be rational is to do as well as possible at minimizing expected incorrectness" (217). These claims seem to entail (together with Wedgwood's semantics for 'ought') that there is, after all, a subjective 'ought' that coincides with rational requirements – namely one that relates to the value of correctness (rather than the value of rationality) and a probability distribution that

represents the agent's epistemic perspective (rather than the omniscient perspective). It is puzzling that Wedgwood does not mention this himself. Throughout, he speaks of the 'ought' entailed by rational requirements in the singular, and in chapter 6 he identifies this 'ought' with an 'ought' that is related to the value of rationality – an 'ought' which his own account classifies as an objective one. He never states the fact that his theory actually delivers two kinds of 'ought' that are entailed by rational requirements: an objective one that is related to the actual value of rationality, and a subjective one that is related to the expected value of correctness.

The Value of Rationality is an ambitious and thought-provoking book, which will be read with great benefit by anyone interested in the theory of rationality. Wedgwood's conception of rationality as an internal virtue that pursues the external aim of correctness is original and highly instructive. It promises to meet several desiderata of the current debate on rationality, by (i) giving a unified account of epistemic and practical rationality, (ii) telling us something informative about the point of rationality rather than listing a number of intuitive requirements, and (iii) elucidating the sense in which we ought to be rational and are criticizable if we aren't. But Wedgwood's book will not only arouse the interest of those who care about the normative status of rationality. In fact, *The Value of Rationality* covers an impressive range of topics and makes novel and systematically interesting contributions to a number of other relevant philosophical topics, such as, for example, the distinction between reasons of the right and reasons of the wrong kind (chapter 2), the principle that 'ought' implies 'can' (chapter 3), and the general structure of virtues (chapter 6). No doubt this book will have a lasting impact on the philosophical debate, and quite rightly so.

As I already indicated, in my opinion the book also has some shortcomings. One of them is that many of the crucial questions that Wedgwood's theory poses and many of the challenges that it faces are only briefly discussed (if at all), while at the same time some of the chapters of the book (particularly chapters 3–5, all of which have been published as self-standing papers before) contain detailed elaborations of points that seem to be only of secondary importance to the book's central question. To be fair, Wedgwood emphasizes several times that his book "is designed to be the first

instalment of a trilogy" (1), and that his theory "will not be presented in full until the whole trilogy is complete" (5). However, given the passion for detail that Wedgwood exhibits with respect to some other, less central issues, it can at times be frustrating to be put off with respect to some of the most burning questions that his book raises.

For example, Wedgwood's interesting proposal that rationality is a matter of minimizing expected incorrectness faces the challenge to explain how paradigmatic instances of structural irrationality, such as means/end-incoherence or akrasia, constitute a violation of the requirement to minimize expected incorrectness. This is an important challenge, not only because accounting for such cases seems to be an independently plausible test case for any conception of rationality, but also because Wedgwood would be talking past Broome, Kolodny and others if his conception of rationality did not apply to the paradigmatic cases of irrationality with which these authors are concerned. Wedgwood spends less than three pages on this issue (232–4), which do not mention the (arguably most difficult) case of means/end-incoherence. The explanation of akrasia essentially relies on the assumption that akratic agents, while intending some action *A*, "rationally have the maximum degree of confidence" in the proposition that "it is better, all things considered, to choose an alternative *B* rather than to choose *A*" (234). However, it also seems akratic to intend *A* while having a less than fully confident belief in that proposition, or a less than fully rational belief (or both). Indeed, only very exceptional cases of akrasia will satisfy Wedgwood's restrictions, and it is for good reasons that the philosophical literature has focused on a much more general phenomenon. It is thus very natural to ask how Wedgwood's conception of rationality treats the vast majority of cases of akrasia in which the belief is not rationally held with maximum confidence. But Wedgwood skims over this question; he jumps to the conclusion that his characterization of rationality "has no difficulty explaining" the rational requirement that bans akrasia (234).

Another set of questions that the book raises but does not address concerns normative conflicts. According to Wedgwood's conception of normativity, there is a variety of values and an infinite number of senses of 'ought' and 'reason', all of which can issue conflicting directives. For example, the value of rationality can conflict with the value of correctness, or with the value of pleasure, or with other values. Relative to each of these values, there will be objective senses of 'ought' that give incompatible

prescriptions, not to mention the innumerable less objective senses of 'ought'. It looks like this will have devastating effects on practical reasoning, unless there is a privileged notion of 'ought' and 'reason' that determines the correctness conditions for deliberative conclusions. Are there such privileged notions – and how are they related to the rational 'ought'?

Wedgwood is almost completely silent on such questions. There is one brief passage in which he mentions the view that “when virtues [like justice and charity] conflict, they can be weighed against each other, to determine what is the right thing for the agent to do all things considered”, and he maintains that “rationality does not just count as one virtue among many in this way” (201). This strikes me as exactly right, which is one of the reasons why I favor a conception of rationality as the capacity to respond to reasons – the reasons, if you see the need for this qualification, that correspond to the notion of 'ought' pertinent in conclusions of deliberation. It is difficult to see, however, how Wedgwood's account, according to which rationality is a matter of pursuing one particular kind of value *among many others*, can account for this special, non-conflicting status of rationality. What he says is that “the requirements of rationality are *decisive*” (202) with respect to every 'ought' that is “(a) internalist ..., and (b) determined purely by 'reasons of the right kind' for the attitude in question” (201–2). But Wedgwood told us before that 'oughts' aren't really determined by reasons at all, but by values and probabilities (in fact, he even told us that talk of reasons is best avoided). As far as I can see, the values relevant for 'oughts' that correspond to 'reasons of the right kind' seem to be the values of rationality or correctness. So, what Wedgwood tells us is that rational requirements do not conflict with (certain kinds of) 'oughts' that are determined by the value of rationality or the value of correctness. But surely, the requirements of justice or charity also do not conflict with certain kinds of 'oughts' that are determined by the values of justice or charity. It's thus not clear how we made any progress towards understanding the special status of rationality that would explain why rationality isn't simply to be weighed against other values.

In the immediately following passage, Wedgwood seems to suggest that the special normative status of rationality is due to its inescapability as a “constitutive virtue of thought” (202). But it is difficult to see why the descriptive or conceptual fact

(if it is a fact) that “all thinkers have at least some disposition to conform to the most basic requirements of rationality” (202) should be taken to support the assumption that rationality is (in some sense) *normatively* decisive.

A further general worry that one might have with this book is that, at bottom, Wedgwood seems to take for granted his most important claims rather than arguing for them. The central claim that rationality is a value, despite being summarized as a thesis that was “argued” (196) and “defended” (200) in chapter 6, is, as far as I can see, neither argued nor defended in that chapter, but simply postulated. In later chapters, Wedgwood aims to explain the value of rationality by showing how it is related to the value of correctness. But again, the thesis that correctness is a value is simply presupposed as “intuitively clear” (231). Now, there is nothing wrong in principle with starting a philosophical inquiry from assumptions that one takes to be intuitively plausible, but it seems that the claims that rationality and correctness are non-instrumentally valuable are in fact very controversial, especially in the dialectical context at issue. For example, it seems to me far from obvious that the mere correctness of a belief in an utterly trivial truth makes belief in that proposition “genuinely valuable” in a “non-relative or absolute sense”. Moreover, if correctness were a genuine absolute value of this sort, this would seem to provide practical reasons to change the facts in such ways that they make our beliefs true, no less than reasons to adjust our beliefs to the facts. Intuitively, however, there are no such practical reasons to change the facts in order to make our beliefs correct. Perhaps worries like these can be resolved. But in order to be resolved, they need to be addressed. I would have liked to see Wedgwood do more to substantiate his controversial axiological assumptions and defend them in light of such potential problems.*

BENJAMIN KIESEWETTER

Humboldt University of Berlin

* I would like to thank Jan Gertken, Felix Koch, Andy Mueller, Thomas Schmidt, Sebastian Schmidt, Jonathan Way and Ralph Wedgwood for comments on an earlier draft. Work on this review was supported by the DFG-project “Principles of the Deliberative Ought”.

References

- Broome, John. 2013. *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.
- Kiesewetter, Benjamin. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.
- Kolodny, Niko. 2005. "Why Be Rational?" *Mind* 114 (455): 509–63.
- Wedgwood, Ralph. 2017. *The Value of Rationality*. Oxford: Oxford University Press.